

複数視点映像における被写体の姿勢変動を考慮した見え方学習法

田中 秀典^{1) 2)} 北原 格^{1) 3)} 斎藤 英雄^{1) 2)} 村瀬 洋^{1) 4)} 小暮 潔¹⁾ 萩田 紀博¹⁾

- 1) ATR メディア情報科学研究所 〒619-0288 京都府けいはんな学研都市光台 2-2-2
- 2) 慶應義塾大学 大学院理工学研究科 〒223-8522 横浜市港北区日吉 3-14-1
- 3) 筑波大学 大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1
- 4) 名古屋大学 大学院情報科学研究科 〒464-8603 名古屋市千種区不老町

E-mail: hiddenori@atr.jp

あらまし 複数視点映像を用いて被写体の見え方を動的に学習する手法について述べる。個人識別を目的とした見え方の学習においては、密に配置されたカメラから得られた様々な方向の見え方を用いるのが一般的である。しかし、実世界を撮影対象とした場合、常にカメラを密に配置できるとは限らない。本研究では、十分な枚数の見え方が撮影されていない初期の段階では、標準顔モデルを介して被写体の見え方を補間した結果を用いて固有空間の生成を行う。その後は、新たな映像が追加撮影される度に、カメラと被写体との相対的な姿勢を推定し、標準顔モデルの見え方を更新することにより、固有空間の識別能力を向上させる。監視カメラで撮影した映像に対して本手法を適用する実験により、提案手法の有効性を確認する。

キーワード 3次元形状推定, パラメトリック固有空間法, 見え方の学習, 見え方の補間, 監視カメラ

Dynamically Visual Learning with Sparsely Distributed Cameras

— Consideration on Postural Change of a Subject —

Hidenori Tanaka^{1) 2)} Itaru Kitahara^{1) 3)} Hideo Saito^{1) 2)} Hiroshi Murase^{1) 4)}
Kiyoshi Kogure¹⁾ and Norihiro Hagita¹⁾

- 1) ATR 2-2-2 Hikaridai, Keihanna Science City, Kyoto, 619-0288 Japan
- 2) Keio University 3-14-1 Hiyoshi, Kouhoku-ku, Yokohama, Kanagawa, 223-8522 Japan
- 3) University of Tsukuba 1-1-1 Tennoudai, Tsukuba Science City, Ibaraki, 305-8573 Japan
- 4) Nagoya University Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8603 Japan

E-mail: hiddenori@atr.jp

Abstract We propose a dynamic visual learning method, which aims to identify people by using sparsely distributed multiple surveillance cameras. In learning method for personal identification, many images captured by densely distributed multiple cameras are generally demanded. However, a practical setup of surveillance cameras does not have dense distribution. In the proposed method, densely distributed multiple images are virtually captured by interpolating the appearance in the sparsely distributed images with a simple 3D shape model, and generate an initial eigenspace in the initial learning step. When another image is captured, the image is projected onto the 3D shape model as texture information to improve the appearance of the interpolated multiple images, and we regenerate the eigenspace of the updated multiple images. The discernment capability for personal identification of the proposed method is shown by experimental results.

Keyword 3D shape model, parametric eigenspace, visual learning, view interpolation, surveillance camera

1. はじめに

近年盛んに報道されている治安悪化の問題を解決する手段として、バイオメトリクス認証技術に注目が集まっている[1][2]。中でも、街の彼方此方に設置された監視カメラを活用して、安心・安全な暮らしをサポートしてくれる技術の開発は、今後さらに重要性を増すであろう。このような“見守りサービス”を実現する際には、パッシブな映像センサによって人物を識別する技術が、大きく貢献すると考えられる。

個人識別を目的とした研究では、密に配置されたカメラによって撮影された多方向からの見え方画像を用いて学習を行うのが一般的である[3]。しかし、実空間を撮影対象とした場合、このような画像が常に獲得できるとは考え難い。一方で、街の彼方此方に監視カメ

ラが設置された環境中を被写体が動き回る場合、時間・空間的に不連続ではあるが、様々な方向からの見え方画像を獲得することが可能である。我々は、実空間における個人識別技術の実現には、このような時間・空間的に離散的な多視点画像を効率良く選択・学習が可能で個人識別手法の開発が必要であると考えている。

本稿では、環境中にまばらに配置された複数台のカメラによって撮影された画像を用いて、個人識別のための被写体の見え方画像を動的に学習する手法について述べる。以下、第2章では従来研究、第3章では提案手法の概要を述べる。第4、5章では手法の具体的な説明を行い、第6章では実験とその結果を述べ、提案手法の有効性を示す。

2. 顔画像を用いた個人識別に関する研究

人間が個人を識別する際には、まず相手の顔を見るように、人間の顔の見え方は、個人認証にとって重要な情報の一つであり、実際に顔の見え方情報を用いた個人認証システムの開発が活発に行われている[4][5]。このような認証システムでは、真正面、または、ほぼ正面の見え方を入力情報として要求し、識別プロセスでは局所的な情報（例えば目、鼻および口）を抽出し、識別を行うものが多い。ドアロックの開錠処理のように、被写体が積極的にカメラに向かい合うような状況では、良好な入力情報が獲得可能であり、緻密な学習と高い識別精度が実現されている[6]が、環境中に設定された監視カメラを用いて見え方情報の獲得を行う場合には、被写体の位置・姿勢をシステムの要求通りに制御することは、現実的であるとは言いがたい。むしろ、そのような状況下では、システムが要求する見え方の獲得は困難であると考えるのが自然であろう。

パラメトリック固有空間法は、様々な方向の見え方を学習し、その結果生成される固有空間中において物体の識別を行う手法である[7]。パラメトリック固有空間法では、画像そのものを特徴量として識別処理を行うため、局所的な情報の抽出精度の影響を受けにくく、より実空間での利用に適した手法である。しかし、様々な姿勢で写りこむ被写体に対して安定した識別を行うためには、事前に被写体の様々な見え方を学習しておく必要がある。実空間にそのような研究室の撮影スタジオと同等の環境を構築することや、不特定多数の来訪者全てに対して緻密な見え方の学習を行うことは現実的ではない。

Muraseらは、少数の見え方情報を用いて粗く生成した固有空間をキュービック・スプライン関数で補間することにより、画像間の見え方の変化を補おうとしている[8]。しかしながら、本研究で対象とする画像間の見え方の変化は、Muraseらが対象としたものよりもはるかに大きい。撮影された被写体の3次元的な見え方の変化が考慮されないキュービック・スプライン関数では、正確な補間が困難である（6.1節において再度、見え方情報の補間法に関する考察を行う）。

自由視点映像生成技術は、複数台のカメラにより撮影された画像を用いて、対象空間の3次元形状の復元を行い、被写体の3次元的な変化を考慮した任意視点位置からの見え方を再現する手法である[9]。本研究では、自由視点映像生成技術を用いて、少数枚のまばらな見え方を補間することにより、被写体のまわりに密にカメラを設置した状況を仮想的に再現し、上述した問題の解決を試みる。

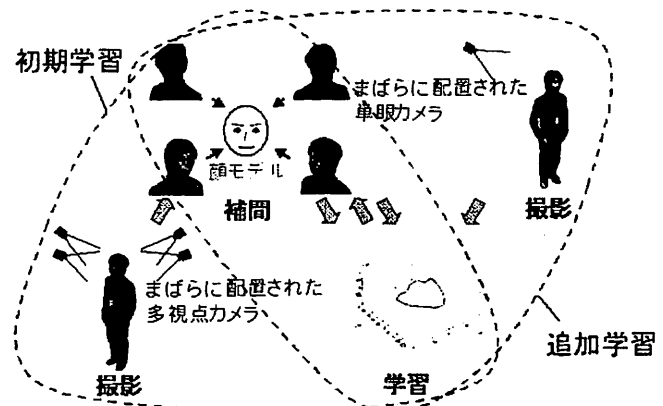


図1：複数視点映像における被写体の姿勢変動を考慮した見え方学習法

3. 複数視点映像における被写体の姿勢変動を考慮した見え方学習法

提案する学習手法は、図1に示すように2段階の処理により構成される。以降、第1段階を「初期学習」第2段階を「追加学習」と呼ぶこととする。

3.1. 初期学習

3次元形状モデルを介して得られた補間画像を用いて、初期の固有空間を作成する。まず、Shape form Silhouette法を用いて被写体のVisual Hullを復元する[10]。そして、そのVisual Hullを覆うように3次元顔モデルをフィッティングすることで、被写体の3次元顔形状の復元を行う。復元モデルを用いて3次元的な見え方の変化を再現した補間処理を行うことにより、仮想的に被写体のまわりに密に監視カメラを設置した状況を再現し、初期の固有空間を生成する。

3.2. 追加学習

初期学習で生成した固有空間を動的に更新し、固有空間の個人識別能力の向上を行う。まず、初期学習で学習済みの人物の追加撮影を行う。その入力画像を初期学習で生成した固有空間に投影し、固有空間中での距離が最も小さい補間画像を探索する。これにより、カメラと被写体の相対的な姿勢を粗い精度で推定する“グローバル探索”を行う。監視カメラによる撮影時には、被写体の姿勢を完全に制御することが困難であるため、グローバル探索で推定された姿勢と、初期学習段階で想定していた被写体の姿勢との間には、若干の差異が生じる可能性がある。そこで、その差を吸収するために、グローバル探索で得られた姿勢の近傍で3次元モデルの姿勢をわずかに変化させた画像を生成し、その画像群と入力画像とのマッチング処理により、さらに詳細なカメラと被写体の相対的な姿勢を探索する“ローカル探索”を行う。

推定されたカメラパラメータを用いて入力画像を3次元モデルに投影することにより、より実際の見え方の変化に忠実な補間画像を生成する。この新しい補間画像を用いて固有空間を再構成する処理を繰り返すことにより、固有空間の個人識別能力を向上させる。

4. 初期学習

本章では、まばらに配置されたカメラの画像から3次元形状モデルを介して初期の固有空間を生成する手法について述べる。

4.1. 頭部領域の切り出し

実空間において見え方の学習を行う場合、被写体を同一箇所に長時間停留させることが困難であるため、できるだけ短時間で学習に必要な見え方情報を獲得する必要がある。頭部領域の切り出しには様々な手法が存在するが、我々は、画像中に被写体の全身が写っている状況を想定し、頭部領域を切り出す手法を提案する。

Koyamaらは、全ての物体が高さ Y の平面上に存在すると仮定することにより、式(1)に示す2次元射影変換 \mathbf{H} を求め、頭上カメラ中の観測位置から物体の3次元位置を実時間で推定している[11]。しかし、この手法は、高さが変化する物体に対応しておらず、また、実際には仮定平面上に存在しない点を天井画像中での物体の観測点 (u, v) としてしまった場合、推定3次元位置に誤差が生じるという問題がある。

$$\lambda \begin{bmatrix} X & Z & 1 \end{bmatrix}^T = \mathbf{H} \begin{bmatrix} u & v & 1 \end{bmatrix}^T \quad (1)$$

そこで我々は、この2次元射影変換を用いた物体の3次元位置推定手法を、対象物体の高さの変化に対応可能となるように改良した[12]。まず、図2に示すように、地面(高さ0)と高さ Y_h の2枚の平面を空間中に設定し、頭上から撮影された画像と各平面の間の2次元射影変換行列 $\mathbf{H}_0, \mathbf{H}_1$ を求める。任意の高さ Y_n に設定した平面の2次元射影変換行列 \mathbf{H}_n は、式(2)に示すように $\mathbf{H}_0, \mathbf{H}_1$ の内挿処理によって求めることができる。

$$\mathbf{H}_n = ((Y_h - Y_n)\mathbf{H}_0 - (Y_n)\mathbf{H}_1) / Y_h \quad (2)$$

平面の高さ Y_n を0から Y_h の間で変化させながら、頭上画像から抽出した前景領域を平面上に投影する。対象物体が人間のように地面に直立している場合、実際に物体が存在している位置 (X, Z) では、平面の高さに関係なく前景領域が投影されるはずである。論理積などを用いて複数の写像を統合することにより、この物体の位置 (X, Z) を推定する。本手法によれば、上方への処理において、最後に写像が存在した平面の高さを求めることにより、対象物体の存在する位置と同時に高さの推定も可能となる。

推定された被写体の顔の3次元位置の周辺に適当な大きさのバウンディングボックスを配置し¹、このバウンディングボックスの8頂点を式(3)により撮影画像上に投影し、それらの点同士を結んだ凸領域を頭部領域とする。

$$\lambda \begin{bmatrix} u_n & v_n & 1 \end{bmatrix}^T = \mathbf{P} \begin{bmatrix} X_n & Y_n & Z_n & 1 \end{bmatrix}^T \quad (3)$$

¹ バウンディングボックスの大きさは標準的な人の頭部サイズに設定するのが妥当であろう。

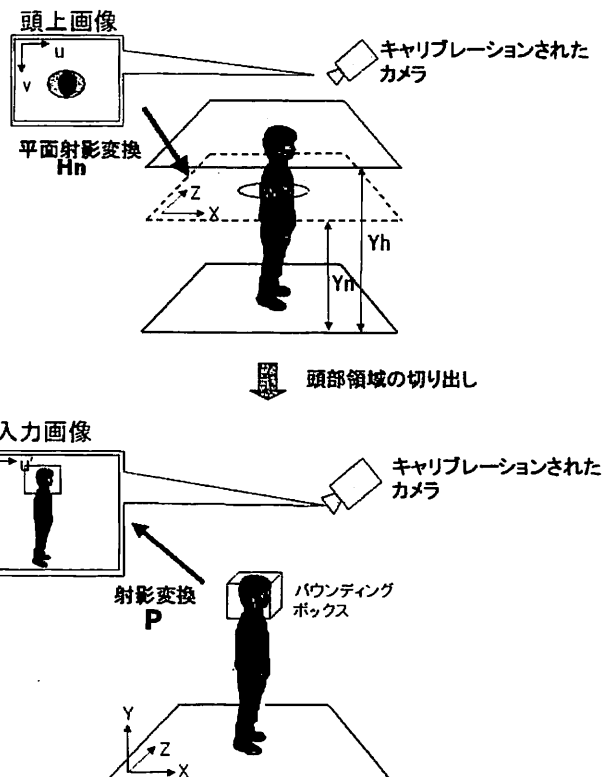


図2：高さの変化に対応可能な頭部領域の切り出し

4.2. 3次元形状の推定と見え方の補間

多視点画像が獲得されている場合、それらに Shape from silhouette 法を適用することで、被写体の3次元形状を復元することが可能である。しかし、人物の顔は起伏に富んでおり、本研究で想定しているように、まばらに配置されたカメラによる少数枚の画像しか獲得されていない場合、被写体の実際の3次元形状と復元された3次元形状の間には大きな誤差が存在する。一方で3次元形状の正確さは補間画像の質に大きく影響する。本研究では、図3に示す Galatea プロジェクト[13]が提供する3次元顔モデルを使用することにより、少数枚の画像からでも、顔の凹凸の影響を反映した、ある程度正確な見え方の変化を実現する。

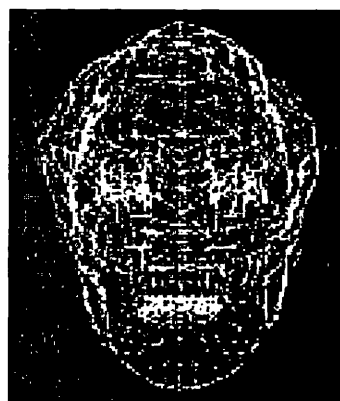


図3：Galatea プロジェクト・3次元顔モデル[13]

被写体毎の個性により顔の大きさや位置が異なるため、3次元顔モデルの大きさ・位置を適切に設定する必要がある。Shape form Silhouette 法により得られるおおまかな形状情報を用いて、顔の大きさ・位置の推定を行う。まず、図4に示すように、地面に平行な高さ y の平面 $p(y)$ 上で被写体の頭部領域の断面を推定し [14]、それに円をフィッティングすることで円の中心 $c(y)=(X,Y,Z)$ と半径 $r(y)$ を求める。頭部領域全体をカバーするように高さ y を変えながら同様の処理を行うことで、複数の円筒により頭部形状が復元される。復元された円筒のうち、ゼロでない最小の半径を持つ円筒の高さと最も高い円筒間の距離を顔の高さ l とする。この顔の高さ l を用いて3次元顔モデルの大きさを調整することにより、顔の大きさの個人差を反映する。次に、復元形状の中心座標を求め、それを3次元顔モデルの原点とすることより適切な位置に3次元顔モデルを設定する。

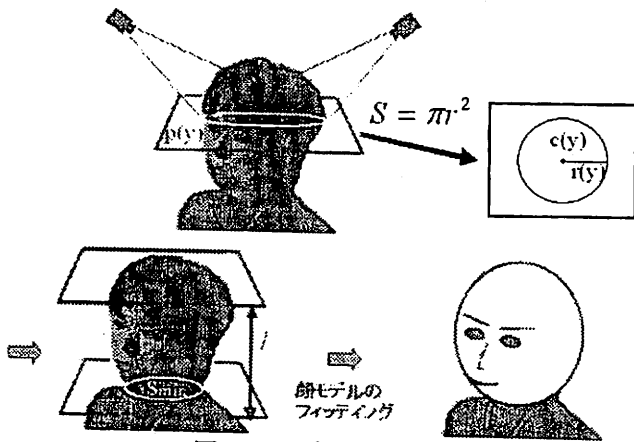


図4：3次元形状の推定

複数台のカメラによって撮影された画像をプロジェクティブ・テクスチャマッピング [15] を用いて、推定3次元顔モデル上にマッピングすることで、3次元形状モデルに見え方情報を付加する。このとき、仮想カメラから入力カメラまでの距離に応じたテクスチャブレンド処理を行う。3次元形状モデルの周囲に適当な間隔で仮想カメラを配置し、レンダリング処理によって補間画像を生成する。

4.3. 固有空間の作成

主成分分析 (PCA) は古くから顔認識システムで用いられてきた手法である [16]。本章では、簡潔に PCA を用いたパラメトリック固有空間法について述べる。

4.3.1. 正規化

固有空間法を用いるためには、画像がすべて同じサイズである必要がある。また、照明の強さ、あるいはカメラの露光設定の差異による変化が、画像に影響を与えないことが望ましい。本手法では、固有空間生成の前処理として“大きさ”と“明るさ”各々について正規化処理を行う。

大きさの正規化では、抽出された頭部領域が、頭頂部を中心とした 128×128 ピクセルに収まるような拡大縮小処理を施す。明るさの正規化では、各画像 \hat{x}_i (画像のラスラスキャンを行い、画素値を要素としたベク

トル) が、式 (4) で示されるようにノルムが 1 になるように正規化される。

$$x_i = \hat{x}_i / \|\hat{x}_i\| \quad (4)$$

本手法は、位置・姿勢・照明条件が完全に制御可能な CG 空間を介して学習用画像を生成するため、固有空間の精度を大きく左右する正規化を性格に実現可能であるという特長を有する。

4.3.2. 初期固有空間の作成

正規化後の画像を用いて主成分分析により固有空間を作成する。まず、画像ベクトル集合の平均ベクトル c を求め式 (5) により計算されたベクトル X から共分散行列 Q を求め固有方程式 (6) を解くことで固有値 i の大きい順に固有ベクトル e_i を算出する。なお、ここで N は学習画像の数である。そして累積寄与率により固有空間を表現する次元 k を決定する。

$$X = [x_1 - c, x_2 - c, \dots, x_N - c] \quad (5)$$

$$\lambda_i e_i = Q e_i \quad (Q = X X^T) \quad (6)$$

式 (7) により、各画像を姿勢のパラメーター (パン方向に 1 度ずつ全周で 360 パラメーター) と関連付けて固有空間に投影する。その結果、1 枚の画像は固有空間の中では、一つの点としてプロットされる。上記のように小刻みに姿勢を変化させた場合、連続する 2 視点間での姿勢変動は小さくなるため、見え方の相関性が強く、固有空間の中では互いに接近するため、全補間画像を固有空間に投影すると、連続的なループを形成する。

$$g_i = [e_1, e_2, \dots, e_k]^T (x_i - c) \quad (7)$$

5. 追加学習

本章では、初期学習時に用いたカメラ以外のカメラで撮影された映像を利用して、初期学習で生成された固有空間を動的に更新する手法について述べる。

5.1. グローバル探索

新しく入力画像が撮影された場合、4.1 節で述べた処理により頭部領域を切り出す。切り出し領域に対して正規化処理を行った後、4.3.2 節で求めた固有ベクトルと平均ベクトルを算出し、式 (8) により固有空間中に投影する。

$$z_j = [e_1, e_2, \dots, e_k]^T (y_j - c) \quad (8)$$

なおここで y_j は入力画像の画素値を要素としたベクトルである。入力画像に最も見え方が近い補間画像を探索するために、固有空間上でのユークリッド距離を計算する。補間画像生成時に関連づけておいた仮想カメラの位置・姿勢パラメーターを参照することにより、カメラと被写体の相対的な姿勢を粗い精度で推定する。

5.2. ローカル探索

実空間中を移動する人の姿勢は制御が困難であるため、新しく撮影された入力画像は、初期学習段階の姿勢とは異なる可能性がある。本節では、初期学習時の姿勢と追加学習時の姿勢の誤差を吸収するローカル探索について詳しく述べる。

5.2.1. 合成画像の生成

推定した3次元顔モデルを用いることにより、様々なカメラ位置から顔の合成画像を生成することが可能である。グローバル探索で一致した画像と被写体の姿勢が少しだけ異なる画像を生成するために、3次元顔モデルを、x、yおよびz軸（x軸は首、y軸は頭、また、z軸はx軸とy軸に垂直）のまわりに微小角度だけ回転させた合成画像を生成する。例えば、3次元顔モデルをx、yおよびz軸のまわりにそれぞれ±30度、±5度、±10度、と5度間隔で変化させると合計195枚の合成画像が生成される。生成された合成画像の例図5に示す。なおここでは、3次元顔モデルが空間中に固定され、それに対して相対的にカメラが移動した位置・姿勢を合成画像生成時のカメラパラメータとして画像に関連づけておく。

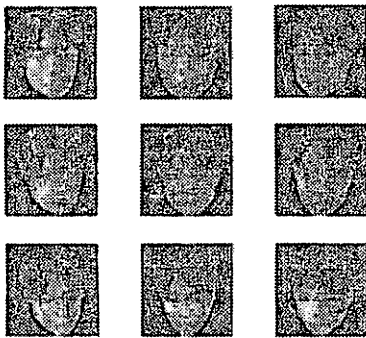


図5：合成画像の例

5.2.2. 合成画像とのマッチング

式(9)で求められる入力画像Iと合成画像S間のSAD (Sum of Absolute Difference)により2枚の画像の類似度を評価する。このとき、 $I(i, j)$ 、 $S(i, j)$ は、入力画像および合成画像の画素 (i, j) での画素値であるとする。Mの値を最小にする画像が最も入力画像に見え方が近い合成画像となる。

$$M(s, t) = \sum_{i, j} |I(i, j) - S(i+s, j+t)| \quad (9)$$

しかし、本研究では最も入力画像に近い合成画像を得ることが目的ではなく、その合成画像を生成したときのカメラパラメータを得ることが目的である。そこで、式(9)を最小にするカメラパラメータを推定するために、多次元の最適化手法である滑降シンプレックス法 [17]を用いる。多次元の最適化では局所解に陥る可能性があるという問題点があるが、それを回避するために、乱数から生成された値を初期値とし、収束させるといった処理を繰り返し行い、得られた収束値の中で最小となるものを入力画像のカメラパラメータとする。

5.3. 固有空間の更新

得られたカメラパラメータを用いて入力画像を3次元形状モデルに投影することにより、被写体の3次元形状モデルの見え方が更新される。その後、4.2節と同様の処理で、3次元形状モデルのまわりに仮想カメラを回転させることにより補間画像が生成される。そして、それらを用いて固有空間を再構成する。この追加学習を繰り返し行うことで、固有空間の個人識別に対する識別能力が向上するものと考えられる。

5.4. 姿勢変動による見え方変動の影響を除去した識別手法

前節で作成した固有空間には、姿勢変動による見え方の変化と、個人差による見え方の変化が混在している。一般的に、個人性よりも姿勢変動の方が見え方の変化が大きいため、識別処理にもその影響が強くなってしまおうという問題があった。識別処理を、姿勢変動識別と個人識別の2段階に分割することにより、この問題を解決する。

まず、5.1節、5.2節で述べた手法を用いた識別処理を行う。ただしこの場合、“被写体が誰であるか”については考慮せず、単に追加学習で撮影された画像中の被写体と最も近い姿勢を検出する処理のみを行う。次に、検出された姿勢情報と同じ姿勢の複数人物画像群を用いて、新たに固有空間を作成し、その空間中における最近傍探索を行うことにより、個人識別を行う。姿勢変動が主成分に現れる固有空間ではなく、個人性が主成分に現れる固有空間を用いて識別を行うことにより、個人識別の精度向上が実現される。

5.5. 三角形関数による選択的学習

追加学習を繰り返すうちに、それまでに学習した画像と類似した角度から撮影した画像が追加されるケースが生じる。そのような画像に対して追加学習を行い、固有空間を更新する場合、被写体と背景の領域分割時に発生した誤差が、識別率に悪影響を与える可能性がある。また、過去に学習が行われた姿勢の近傍については、十分な見え方情報が獲得されているため、新たに追加学習処理を行うことは、冗長であるといえる。本研究では、この問題を解決するために、選択的に追加学習を行う手法を開発する。

選択的学習の基準として、姿勢をパラメータとした重み関数を用いる。具体的には、補間処理によって生成された画像群と学習済みの入力画像のSAD値を計算し、それらの値を直線補間することにより、図6に示すような三角形関数を作成する。それまでに学習した姿勢の近傍では、小さな重みが与えられているのがわかる。追加学習時には、算出された見え方の類似度に、この関数値を重みとして反映させることにより、その姿勢の見え方画像を追加学習するかどうかを判定する。追加学習が行われるたびに、学習画像のSAD値を用いて、三角形関数を更新し、以降の追加学習処理の重みとする。

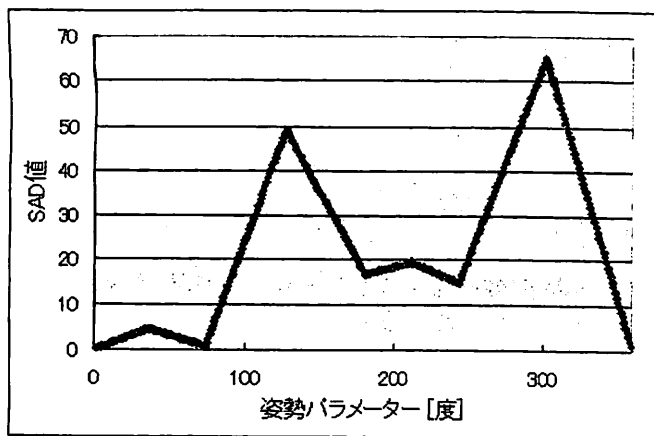


図6：初期学習時（学習枚数4枚：0°，70°，180°，240°）における三角形関数の様子

6. 実験

提案手法の有効性を確かめるための実験を行った。撮影用カメラとしてソニー製 Web カメラ（sony SNC-Z20）を用いた。入力画像の大きさは640×480pixelであり固有空間の次元は累積寄与率が95%となる27次元とした。

6.1. 補間手法の比較

まばらに配置されたカメラから撮影された入力画像に対して、3次元形状モデルを介した補間処理と、キュービック・スプライン関数を用いた補間処理とを比較する実験を行った。このとき、約22.5度間隔（全周16枚）で撮影した実写画像を比較時の基準とした。

図7は、キュービック・スプライン関数を用いて補間した結果と実写画像16枚を比較した結果である。また、図8は、3次元形状モデルを介した補間結果と、実写画像16枚を比較した結果である。図7・図8では、軌跡の特徴の視認性を向上させるために、キュービック・スプライン関数を用いて実写画像の写像の補間を行っている。なおここでは結果を見やすくするために固有空間を同一視点から見た第1主成分から第3主成分を主軸とするグラフで示す。

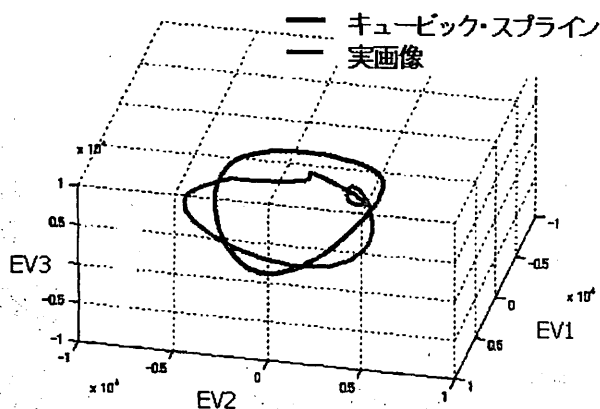


図7：キュービック・スプライン関数による補間

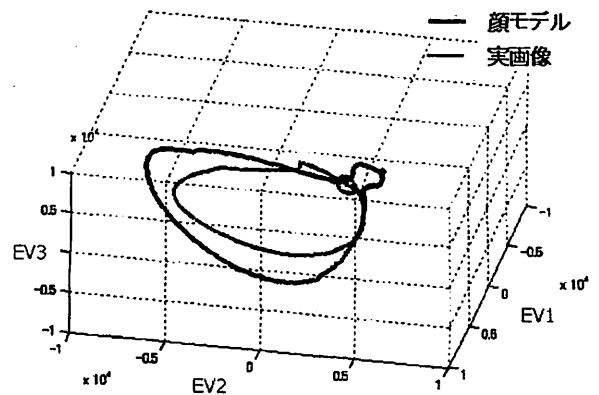


図8：提案手法による補間

図7と図8を見比べてみると、キュービック・スプライン関数による補間よりも、3次元形状モデルを介した補間の方がより実写画像に近い補間がなされていることがわかる。このことから、まばらな入力画像しか与えられない場合には、これまで固有空間の補間処理に用いられてきたキュービック・スプライン関数による補間より、3次元形状モデルを介した補間処理の方が、識別能力の高い固有空間の生成が可能であるといえる。

6.2. 識別能力の検証

初期学習時の補間画像が追加学習によってどのように更新されるかを図9に示す。図9の上段図は初期学習時の補間画像であり、図9の下段図は追加学習により更新された補間画像である。図9の丸の中を見てもわかるように、追加学習により補間画像の見え方が改善されている。結果として、識別能力の高い固有空間が作成できると考えられる。

しかし、図9下段図を見てもわかるように追加学習により補間画像の見え方にノイズが加わる場合がある。これは、カメラパラメータの推定誤差や頭部領域の切り出し時の誤差によってテクスチャマッピングにずれが生じた事が原因と考えられる。

次に、追加学習が進むにつれて、固有空間の識別能力がどのように改善されていくのかを確認するために、4人の人物を被写体とした識別実験を行った。まず4人の内の1人を識別対象に設定し、識別対象について3次元形状モデルを用いた補間画像生成を行う。次に、その画像を用いて生成した固有空間に、4人全員の顔画像50枚を投影し、補間画像によって形成される固有空間中の軌跡と投影結果との距離を算出する。そして、判別対象とそれ以外に最もよく分離する距離（閾値）を求める。閾値設定に用いたのは別の4人の顔画像50枚を再度固有空間に投影し、先程求めた閾値処理により識別対象とそれ以外を識別する実験を行った。図10に追加学習による識別率の変化の様子を示す。図10において追加学習が進むにつれて識別率が向上し、識別能力が向上していることがわかる。しかし、図10の点線において追加画像の枚数が3枚から4枚になるときに識別率が低下している。これは、先ほど述べた追加学習時のテクスチャマッピングのずれや切り出し誤差による影響によるものと考えられる。

そこで、テクスチャマッピングのずれを改善するために目鼻口の特徴点を基準とし、手動でカメラパラメータの補正を行った。手動で補正処理を行った場合の追加学習による識別率の変化の様子を図 10 の実線グラフとして示す。追加画像の枚数が 3 枚から 4 枚になるときの識別率の低下が改善されていることが確認できる。このことにより、カメラパラメータが正確に求められれば、本手法で提案する見え方の動的学習の仕組みが有効に機能することが示されている。

Initial Images



Updated Images



(a) PersonA

Initial Images



Updated Images



(b) PersonB

Initial Images



Updated Images



(c) PersonC

図 9：追加学習の様子

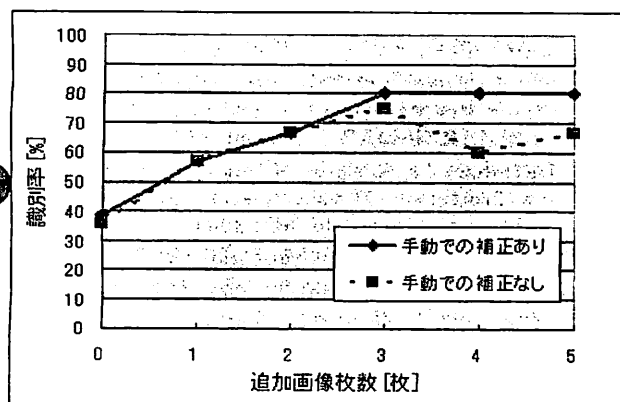


図 10：追加画像の枚数と識別率

6.3. 識別手法の改良と検証

前節で紹介した実験結果では、被写体の姿勢変動の影響が大きいため、識別率が約 80% 以上に向上することがなかった。そこで、5.4 節で紹介したように、姿勢の変動の影響を取り除いた後に、個人識別する手法を適用した実験を行う。まず、先ほどと同様に 4 人の内の 1 人を識別対象に設定し、識別対象について補間画像によって形成される固有空間中の軌跡と入力画像の投影結果との距離を算出する。そして、被写体の姿勢を推定する。その後、同じ姿勢の 4 人全員の画像を用いて固有空間を作成し、そこに再び入力画像の投影

し判別対象とそれ以外に最もよく分離する距離(閾値)を求める。そして、閾値設定に用いたのは別の 4 人の顔画像 50 枚を同様に投影し、先程求めた閾値処理により識別対象とそれ以外を識別する実験を行った。図 11 に追加学習による識別率の変化の様子を示す。図 10 に比べ識別率が向上していることが確認できる。しかし、図 11 の点線において追加画像の枚数が 1 枚から 2 枚になるときに識別率が低下している。これは、先ほどと同じく追加学習時のテクスチャマッピングのずれや切り出し誤差による影響によるものと考えられる。そこで、テクスチャマッピングのずれを改善するために目鼻口の特徴点を基準とし、手動でカメラパラメータを補正した。図 11 の破線に手動による補正後の追加学習による識別率の変化の様子を示す。先ほどの識別率の低下が改善されている。しかし、まだ識別率が若干低下しているところがある。これは、既に学習された画像の姿勢近辺に追加学習されたことによるものと考えられる。

さらなる識別精度の向上を目指し、5.5 節で紹介した姿勢をパラメータとした三角形関数モデルにより追加学習を制御する手法を用いた識別実験を行った。図 11 の実線に三角形関数モデルを適用した場合の追加学習における識別率の変化の様子を示す。識別率が低下することなく追加学習が進むにつれて識別率が約 90% まで向上していることがわかる。

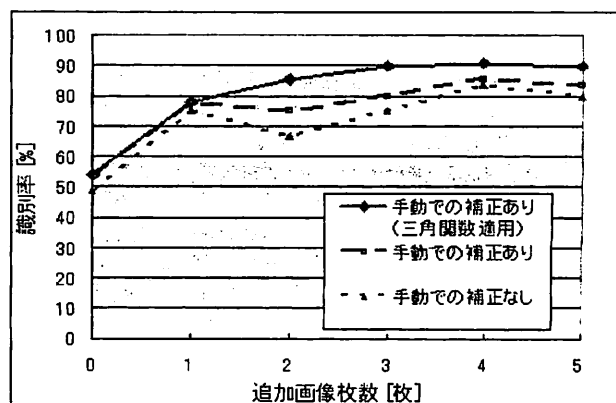


図 11：追加画像の枚数と識別率 (姿勢変動の影響除去後)

6.4. カメラ台数の検討

実空間において、本提案手法による個人識別処理を実現する場合、初期学習用カメラの設置台数とその配置は、非常に重要な検討項目である。そこで、追加学習による識別精度の変化を調査することにより、初期学習時において必要とされるカメラ台数の検討を行う。具体的には、6.3 節で述べた評価法により識別率を算出し、追加学習が行われるにつれて識別率が向上するかどうかにより、初期学習時に設定したカメラ台数の妥当性を評価した。なお、初期学習時に用いるカメラの配置は、図 12 に示すように、カメラ 1 台～6 台とする。被写体の真正面にカメラを設置することは、実空間において現実的ではないため、撮影用カメラは、常に被写体の側面から撮影を行うこととする。

文献

- [1] A.K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, "Biometrics: A Grand Challenge", Proc. of ICPR (ICPR 2004), Vol. 2, pp. 935-942, 2004.
- [2] A. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing", IEEE Trans. on PAMI, Vol. 22, No. 1, pp. 107-118, 2000.
- [3] R. Gross, I. Matthews, and S. Baker, "Appearance-Based Face Recognition and Light-Fields", IEEE Trans. on PAMI, Vol. 26, No. 4, pp. 449-465, 2004
- [4] S. Lao, T. Kozuru, T. Okamoto, T. Yamashita, N. Tabata, M. Kawade. "A fast 360-degree rotation invariant face detection system", ICCV (ICCV 2003), 2003.
- [5] http://www.identix.com/products/pro_security_bn_p_argus.html
- [6] 土居 元紀, 陳 謙, 眞 溪 歩, 大 城 理, 佐 藤 宏 介, 千 原 國 宏, "顔 画 像 照 合 に よ る 解 錠 制 御 シ ス テ ム", 信 学 会 論 文 誌, Vol. J80-D-II, No. 8, pp. 2203-2208, 1997.
- [7] H. Murase, S.K. Nayar, "Parametric Eigenspace Representation for Visual Learning and Recognition", Workshop on Geometric Method in Computer Vision, SPIE, pp. 378-391, 1993.
- [8] H. Murase, S.K. Nayar, "Illumination Planning for Object Recognition Using Parametric Eigenspaces", IEEE Trans. on PAMI, Vol. 16, No. 12, pp. 1219-1227, 1995.
- [9] T. Kanade, P.J. Narayanan, and P.W. Rander. "Virtualized reality: concepts and early results", Proc. of IEEE Workshop on Representation of Visual Scenes, pp. 69-76, 1995.
- [10] A. Laurentini, "The Visual Hull Concept for Silhouette-Based Image Understanding", IEEE Trans. on PAMI, Vol. 16, No. 2, pp. 150-162, 1994.
- [11] T. Koyama, I. Kitahara, Y. Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium", Proc. of ISMAR (ISMAR 2003), pp. 178-187, 2003.
- [12] I. Kitahara, K. Kogure, N. Hagita, "Stealth Vision for Protecting Privacy", Proc. of ICPR (ICPR 2004), Vol. 4, pp. 404-407, 2004.
- [13] <http://hil.t.u-tokyo.ac.jp/~galatea/>
- [14] I. Kitahara, H. Saito, S. Akimichi, T. Ono, Y. Ohta, and T. Kanade, "Large-scale Virtualized Reality", CVPR, Technical Sketches, 2001.
- [15] C. Everitt, "Projective Texture Mapping", White paper, NVIDIA Corporation, 2001.
- [16] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.
- [17] J.A. Nelder and R. Mead, "A Simplex Method for Function Minimization", Computer Journal, Vol. 7, pp. 308-313, 1965.

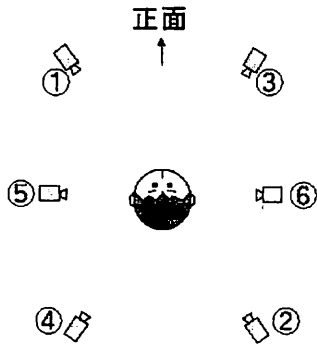


図 1 2 : 初期学習におけるカメラ配置パターン

図 13 に初期学習時におけるカメラ台数と識別率の変化の関係を示す。この図からわかるようにカメラ台数が 1 台～3 台の間は、追加学習が行われても識別率はほとんど変わらず、追加学習の効果がないことがわかる。一方、カメラ台数が 4 台～6 台になると追加学習が行われるほど識別率が向上していることがわかり、追加学習の効果が示されている。視点数の増加に伴い設置の手間もまた増大し、一方で、他の物体による隠れの影響を受けやすくなるなどの問題が発生することから、初期学習におけるカメラ台数は、なるべく少ない方が良く考えられるため、初期学習時におけるカメラ台数は 4 台が適当であることがわかる。

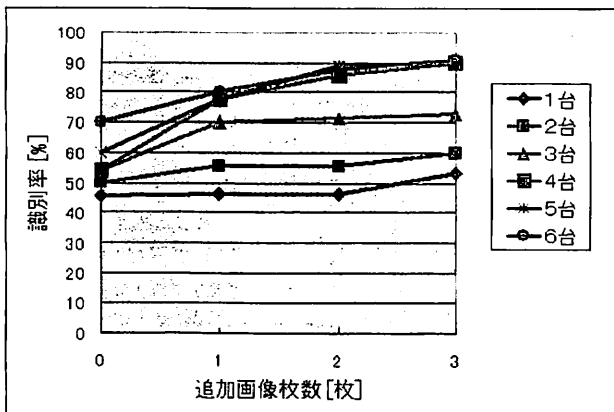


図 1 3 : 初期学習におけるカメラ台数の変化と追加識別率の変化の関係

7. まとめ

まばらに配置された複数カメラからの動的学習法を提案した。3次元形状モデルを介した補間の方が、従来のキュービック・スプライン関数を用いた補間よりも実画像に近いことや、初期固有空間の識別能力が追加学習により向上すること、また、初期学習時における適当なカメラ台数を示した。

今後の課題は、テクスチャマッピングの自動化における精度向上や、撮影増中からの被写体の切り出し誤差の低減などがある。また、より高い識別率のために様々なセンサ情報との統合などについても、今後取り組んでいきたい。

本研究は情報通信研究機構(NICT)の研究委託により実施したものである。