

# 自己連想制約および中間層へのノイズ付加による ニューラルネット識別器の汎化性能向上

## 照明変動下での顔識別の場合

稲吉 宏明<sup>†</sup> 栗田多喜夫<sup>†</sup>

<sup>†</sup>産総研 〒305-8568 茨城県つくば市梅園 1-1-1  
E-mail: †{h.inayoshi,takio-kurita}@aist.go.jp

あらまし 本稿では、汎化性能を向上させるための新たな学習アーキテクチャを提案する。具体的には、通常のニューラルネット識別器の中間層に (1) 入力自己連想を同時に学習させ、さらに (2) 学習過程で中間層各素子に独立なノイズを付加する、という方法を提案し、提案手法の有効性を「照明変動下の顔画像識別問題」で示す。汎化性能の向上には未訓練のデータを「訓練データから補間可能」とすることが重要であるが、提案手法では (1) 自己連想制約により、訓練データの主成分に近い表現を中間層に形成し、さらに (2) 中間層へのノイズ付加により、各主成分への変動が加えられた「訓練データ以外のデータ」も生成して、それらを学習することで補間能力向上が期待できる。  
キーワード 自己連想、ノイズ、照明変動、汎化性、補間、主成分分析

## Adding both auto-association and hidden-layer-noise to neural-network-based-classifiers improves generalization a case for face-classification under variable illumination

Hiroaki INAYOSHI<sup>†</sup> and Takio KURITA<sup>†</sup>

<sup>†</sup> AIST Umesono 1-1-1, Tsukuba-city, Ibaraki, 305-8568, Japan  
E-mail: †{h.inayoshi,takio-kurita}@aist.go.jp

**Abstract** We propose a novel method for learning that improves generalization in neural-network-based-classifiers. The proposed method consists of (1) adding auto-associative-learning and (2) simultaneously adding independent noise to the hidden-layer of the neural-network. We verify this method on the classification problem of faces under variable illumination. Considering the interpolation for untrained samples as key aspect of generalization, we can expect that in our method, neural-classifiers will (1) learn (nearly) principal components of trained samples by auto-association, and will (2) generate and learn the variated samples from trained samples (along the axes of nearly principal components) by added noise, which leads both to increased amount of trained samples and to (hopefully) improved generalization.

**Key words** auto-association, noise, variable illumination generalization, interpolation, PCA

### 1. はじめに

識別器の性能向上には汎化能力が不可欠である。ここで識別器の汎化能力とは、未学習の入力に対しても学習済みの入出力関係に基づいて正しい識別を出力する能力を指す。本稿では、汎化能力向上のための新たな学習アーキテクチャを提案する。

識別器にニューラルネットを利用するという前提で、従来手法 (図 1 左) と提案手法 (図 1 右) の相違点/新規性は (1) 中間層

を共有しての自己連想制約および (2) 中間層各素子への独立なノイズ付加の両方を学習過程に行う点にある。ここで自己連想とはある入力に対して、この入力と同じものを出力することという。

提案手法の意図として、(1) 自己連想制約により入力サンプルの主成分に近い表現を中間層に形成させ、さらに (2) 入力空間にノイズを付加するのではなく、特徴空間である中間層にノイズを付加することにより、特徴空間の各軸に沿った変動を生成

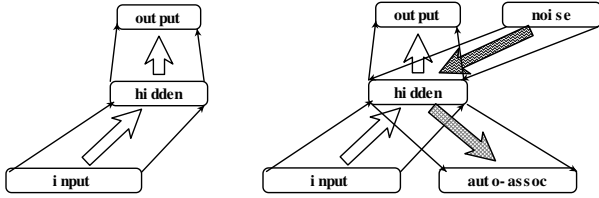


図1 提案手法(右)および従来手法(左)の概略図

し、これを学習させることによる汎化性能向上を期待している。ここで、単純に入力空間にノイズを付加することにより形成されるノイズ後入力と、特徴空間の各軸に沿った変動を入力に付加することにより形成されるノイズ後入力との違いを例えば、入力として画像を対象とした場合の「照明条件の変動により得られる、同一静止被写体の画像集合」で説明する。画像の変動の主成分として「ある固定位置からの光源強度変化」が特徴空間のある軸として学習されるならば、この軸に沿ったノイズによる変動は「固定位置からの光源強度の仮想的変化」に対応する未学習の画像の形成に対応し、同様に、「光源位置を水平平面内で変更した場合の画像の変動」が「特徴空間のある軸の組み合わせからなる空間」として学習されるならば、この軸の張る空間内でのノイズによる変動により「光源位置の水平平面内での仮想的変化」に対応する未学習の画像の形成が期待できる。これらの効果は入力空間に単純にノイズ付加することでは期待できないものである。

なお、高橋らによる文献[1]で自己連想制約を付加する手法は示されており、他方、栗田らによる文献[2]では中間層にノイズを付加する効果が示されているが、両者の協調効果の有効性を示すのは本稿が初である。

汎化性能向上のための関連研究としては、Bishopによる入力へのノイズと正則化に関する研究[3]や赤穂による、学習用データを補間するようなノイズの付加方法の研究[4]、およびMurrayらによる結合加重にノイズを付加した場合の研究[5]が挙げられるが、提案手法は学習用データの変動の主成分に注目していることから、これらの関連研究とは異なる視点に基づいている。

汎化性そのものに関する興味深い研究としては、Tenenbaumらによる研究[6]が挙げられる。彼らは「1つの例からの一般化」についてのShephardの理論を「複数の例からの一般化」へと拡張し、「複数例の分布」に応じて「汎化関数」がいかに変化するかを示している。

本論文の残りの構成は次の通りである：2節で、提案アーキテクチャの学習アルゴリズムを示し、3節で、「照明変動下の顔画像識別問題」を説明する。4節で実験とその結果を示し、5節で結論および今後の課題を述べる。

## 2. 提案する学習アーキテクチャ

本節では、提案アーキテクチャの学習アルゴリズムを示す。

図2に示すような構造のニューラルネット識別器を考える。この主要部分は、 $j$ 番目( $j = 1, 2, \dots, N$ )の入力 $x_j$ ( $M$ 次元ベクトル)を $K$ 個のクラスに識別するために、中間表現 $y_j$ ( $H$

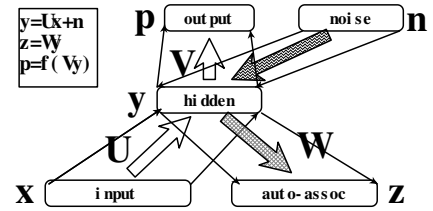


図2 自己連想制約と中間層ノイズ付加のあるニューラルネット識別器

次元ベクトル)を経由して出力 $p_j$ ( $K$ 次元ベクトル)に変換するネットワークである。ただし、中間表現 $y_j$ は、 $z_j$ ( $M$ 次元ベクトル)として入力 $x_j$ の自己連想出力形成にも利用されかつ、 $H$ 個の独立なノイズで構成される $n_j$ ( $H$ 次元ベクトル)から影響を受ける。中間層が線形ニューロンで構成され識別部分に多項logitモデルを用いた場合に、これらの変数間の関係は以下ようになる：

$$y_j = Ux_j + y_0 + n_j \quad (1)$$

$$p_j = f(Vy_j + p_0) \quad (2)$$

$$z_j = Wy_j + z_0 \quad (3)$$

上記で $\{U, VW\}$ は各々、 $H$ 行 $M$ 列、 $K-1$ 行 $H$ 列、 $M$ 行 $H$ 列の行列であり、 $\{y_0, z_0\}$ は各々、 $y_j, z_j$ と同次元のバイアス項( $p_0$ は $K-1$ 次元)である。また、行列 $V$ の各行を1行 $H$ 列ベクトル $v_k$ ( $k = 1, 2, \dots, K-1$ )とし、 $\eta_{jk} = v_k y_j + (p_0)_k$ とした場合、(2)式は次式に示す”softmax”として計算される：

$$p_{jk} = \exp(\eta_{jk}) / \{1 + \sum_{i=1}^{K-1} \exp(\eta_{ji})\} \quad (4)$$

$$p_{jK} = 1 / \{1 + \sum_{i=1}^{K-1} \exp(\eta_{ji})\} \quad (5)$$

((5)式の項を除いた $K-1$ 項の(4)式からなる $K-1$ 次元ベクトルを $r_j$ とする。)

ここで、 $K$ 個のクラス $\{C_1, C_2, \dots, C_K\}$ に識別するための教師信号として、 $K$ 次元の2値ベクトル $\tau = (t_1, t_2, \dots, t_K)^T \in \{0, 1\}^K$ (および $K-1$ 次元の2値ベクトル $t = (t_1, t_2, \dots, t_{K-1})^T \in \{0, 1\}^{K-1}$ )を考える。正解のクラス $C_k$ に対応する要素 $t_k$ のみ1で、他要素は全て0。訓練データの集合を $\{(x_j, \tau_j)\}_{j=1}^N$ で表したとき、訓練データ集合に対する尤度は

$$P(\tau | \mathbf{y}) = \prod_{j=1}^N \prod_{k=1}^K p_{jk}^{\tau_{jk}} \quad (6)$$

となる。

他方、自己連想の学習では入力 $x_j$ とその自己連想出力 $z_j$ の間の誤差 $\epsilon_j^2 = \|x_j - z_j\|^2$ の平均

$$\epsilon^2 = (1/N) \sum_{j=1}^N \|x_j - z_j\|^2 \quad (7)$$

を減らすように学習するが、これは最尤推定の枠組みでは $\sum_{j=1}^N \epsilon_j$ が平均0、分散 $\sigma$ の正規分布に従うと仮定すると

$$L_A = (-1/2) \sum_{j=1}^N \epsilon_j^2 \quad (8)$$

の最大化と等価である。

(6) 式の対数尤度  $L_C = \log P(t|y)$  と自己連想に関する尤度 ((8) 式) の和  $L_C + L_A$  を最大化するパラメータ集合  $\{U, V, W, y_0, z_0, p_0\}$  の更新式 (学習アルゴリズム) は、(9)- (14) 式となる。ただし、学習係数を  $\alpha$  とおき、 $\{x_j, y_j, z_j, t_j, r_j\}$  について各々  $N$  個のベクトルを並べて  $N$  行  $N$  列の行列で表現し ( $x_j \rightarrow X, y_j \rightarrow Y, z_j \rightarrow Z, t_j \rightarrow T, r_j \rightarrow R$ )、さらに  $E = (T - R), F = (X - Z), G = V^T E + W^T F$  とする。

$$V_{new} = V_{old} + \alpha(EY^T) \quad (9)$$

$$W_{new} = W_{old} + \alpha(FY^T) \quad (10)$$

$$U_{new} = U_{old} + \alpha(GX^T) \quad (11)$$

$$(y_0)_{new} = (y_0)_{old} + \alpha(\text{sum}(G)) \quad (12)$$

$$(z_0)_{new} = (z_0)_{old} + \alpha(\text{sum}(F)) \quad (13)$$

$$(p_0)_{new} = (p_0)_{old} + \alpha(\text{sum}(E)) \quad (14)$$

ここで (12), (13), (14) 式の  $\text{sum}(A)$  は行列  $A$  に対して各行の行和 ( $\text{sum}(A)_i = \sum_j (A)_{ij}$ ) からなるベクトルを返す関数である。

### 3. 照明変動下での顔画像識別

本稿の目的が「照明変動下での顔画像識別のための最善手法の提案」ではなく、提案手法の有効性が確認できる例題として「照明変動下での顔画像識別問題」を取り上げていることを前置きとした上で、本節では同問題を説明する。なお、照明変化を伴う顔認識に関しては、岡部らによる、近年の研究動向のまとめ [7] があるため、詳細はそちらを参考されたい。

顔画像の識別において、照明変化による見えの変化は個人差による見えの変化よりも大きいことが指摘されている [8]。“illumination cone” とは、ある物体の任意照明下での画像集合が画像空間内で形成する、原点を頂点とする convex cone を指すが、Belhumeur らは、凸物体かつ Lambert モデルの場合の “illumination cone” が、3 枚の基底で生成可能なことを示した [9]。

照明変動の影響を受けにくい、見え方に基づく認識手法として有名なものは Turk らによる Eigenfaces [10] および Belhumeur らによる Fisherfaces [11] が挙げられるが、我々の把握している範囲で最良の識別性能を示している手法は、Savvides らによる “Corefaces” [12] であり、ほぼ 100 % に近い識別性能を示している。

## 4. 識別実験

### 4.1 識別対象

提案手法の有効性確認のための実験として、顔画像データベース Yale Face Database B [13] を用いた識別実験を行った。

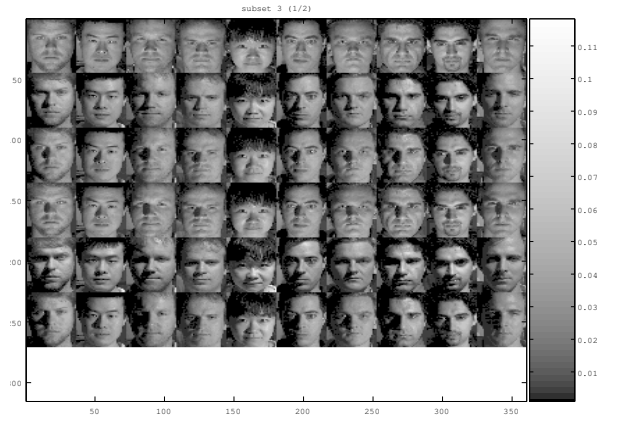


図 3 識別対象の顔画像データサンプル

上記データベースは 10 名の顔を 9 通りの姿勢で各々 65 通りの照明条件下で撮影した  $10 \times 9 \times 65 = 5850$  枚の画像で構成されており、このうち正面向きの 650 枚から  $240 \times 300$  pixel で顔部分を切り出しこれを  $36 \times 45$  pixel にリサイズした画像を実験に用いた。図 3 に 10 名のサンプル画像を示す。同図の各行は共通の照明条件での画像であり、本稿ではこの共通の照明条件下の 10 枚を 1 組と呼ぶ。

ただし、入力画像を  $36 \times 45 = 1620$  次元の画素値ベクトル  $x_{raw}$  としたとき、

$$x = x_{raw} / \|x_{raw}\|$$

で各画像を正規化した。

### 4.2 実験設定

識別器として、中間層の素子数をパラメータとした数通りの

- 従来手法
- 提案手法 = 従来手法 + 自己連想 + 中間層ノイズ付加

の各々で共通の学習係数および反復回数で学習を行い、汎化性能をテストした。ただし提案手法に関しては、付加する一様ノイズ (区間  $[-s/2, +s/2]$ ) のスケール  $s$  もパラメータとしてさらに数通り変化させた。  $s \in \{0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0\}$

学習およびテストは総計 65 組 (650 枚) のサンプルに対して、

- (1) 10 組 (100 枚) で訓練し残り 55 組 (550 枚) でテスト
- (2) 20 組 (200 枚) で訓練し残り 45 組 (450 枚) でテスト

の 2 通りで行った。

### 4.3 実験結果

65 組のサンプル集合を、訓練用+テスト用にランダムに分割する試行を (10+55) 分割と (20+45) 分割について各 100 回行い、各試行で学習後テスト用顔画像 550 or 450 枚が 10 名中の誰であるかを識別した際の誤答数を図 4,5,6 左側に示す。これらの図は中間層サイズが順に 20,10,3 の場合に対応している。同図各々、上段は 550 枚の、下段は 450 枚の場合で、横軸はいずれもノイズスケール  $s$  であり、縦軸は誤答数である。ただし、横軸については、左端の値  $= -1$  の箇所は従来法であり、棒グラフの隣接する左が 100 試行でのメディアン、右が平均値を示す。

さらに、図 4,5,6 の各右側には「誤答数比」すなわち、「従来

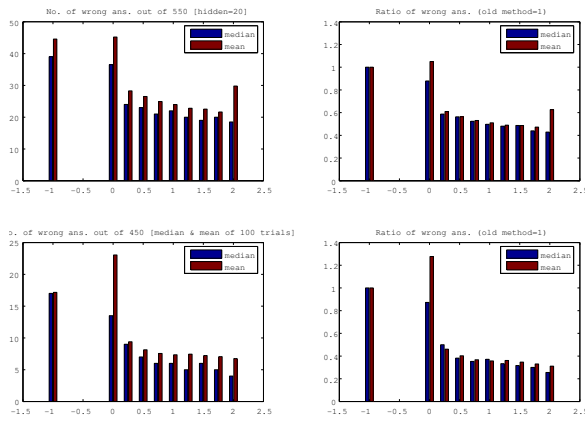


図 4 中間層 20 の場合の誤答数 (左) および誤答数比 (右)

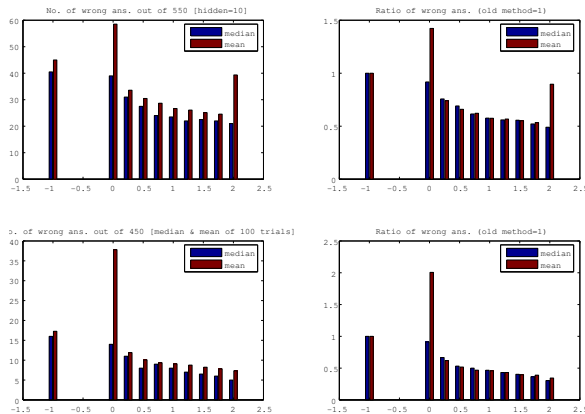


図 5 中間層 10 の場合の誤答数 (左) および誤答数比 (右)

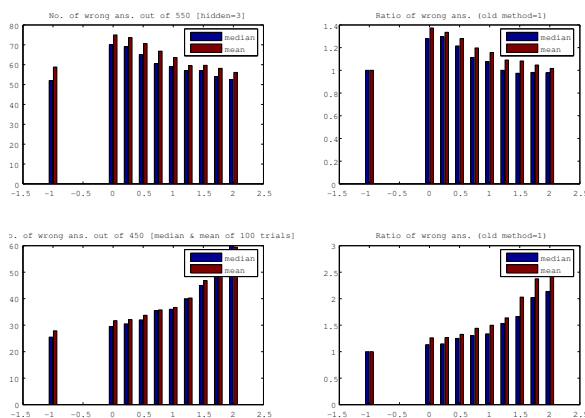


図 6 中間層 3 の場合の誤答数 (左) および誤答数比 (右)

法での誤答数を 1 とした場合の、提案法での誤答数」を示す。横軸および、棒グラフの隣接する左右については、図 4,5,6 の左側と右側で共通である。

一連の図で「誤答数 (左側)」と「誤答数比 (右側)」は単一の試行では「単純なスケール変換の関係」にあるが、複数試行で

統計処理したものは「単純なスケール変換の関係」ではない点に注意されたい。図中、平均値で大きめの値となっている部分は、学習が発散してしまった場合の高い誤答数の影響を受けたためである。発散しない場合は、訓練サンプルの誤答数は総てゼロである。また、ノイズスケールが 2.5 以上になると発散頻度が増加し、誤答数も増大するため図のプロットから除外した。

図 4,5 から、中間層サイズが 3 の場合を除き、適切なスケールのノイズを用いた提案手法により、従来手法の誤答数をおよそ半減できたことがわかる。

なお、図 6 に示されるように、中間層のサイズが 3 の場合提案手法での汎化性能の向上が見られないのは、illumination cone に最少で 3 つの基底が求められており、この空間での変動のみで補間を行うのは困難なためと考えられる。

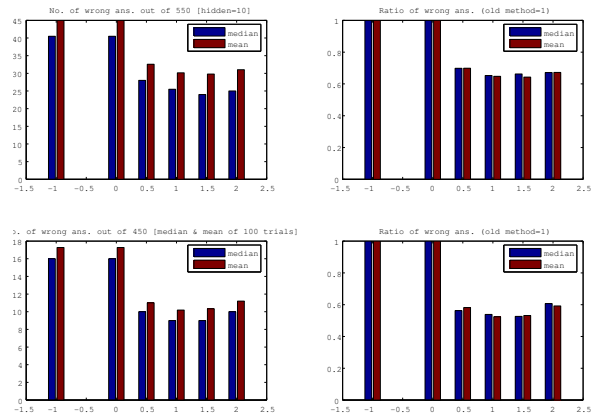


図 7 自己連想無しでの中間層 10 の場合の誤答数 (左) および誤答数比 (右)

対照実験として、「中間層ノイズ付加のみ」で自己連想をつけなかった場合の結果を中間層サイズ 10 の場合のみであるが、図 7 に示す。この図の横軸縦軸等は図 4,5,6 と共通だが、左端はノイズスケール 0 と同一である。図 7 と図 5 の誤答数比の比較により、少なくともこの実験条件の場合には「中間層ノイズ付加のみ」よりもこれに自己連想を同時付加するほうが、汎化性が向上していることがわかる。

参考として、学習結果を表示したものを図 8 に示す。同図は中間層数 10 のときのある同一試行での従来法 (誤答数 34) および提案法 ( $s = 1.75$  で誤答数 1) での学習結果例であり、4 つの行は上から順に以下を表示している：

- (1) 従来法の行列  $U$
- (2) 提案法の行列  $U$
- (3) 提案法の行列  $W$
- (4) 提案法の行列  $WV^T$  と  $z_0$

さらに、図 8 に示した提案法での学習結果に基づいた自己連想による再構成出力例を図 9 に示す。この図は、上 2 列が中間層内の 10 個の全素子に  $[-0.5, 0.5]$  の区間の独立な一様乱数ノイズを 10 通り付加した場合、下 2 列が中間層内の 10 個の各素子ひとつのみに順に大きき 1 のノイズを付加した場合に対応している。ただし、上 2 列の左端の 2 枚は、入力用の「同一人物



図 8 中間層数 10 のときの従来法 (1 行目) および提案法 (2,3,4 行目) での学習結果例

の 65 通りの照明条件に関する平均画像」(1 行目)、および中間層にノイズを付加しない場合の自己連想出力 (2 行目) である。

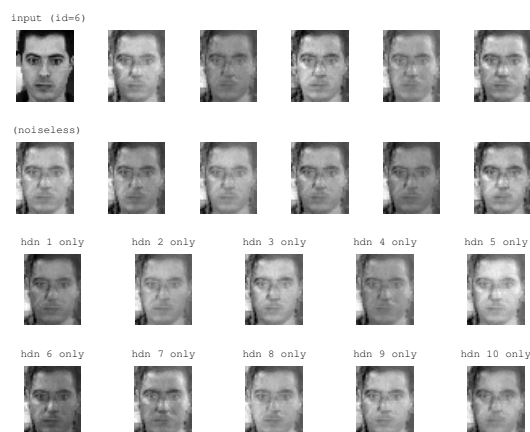


図 9 中間層数 10 のときの提案法での学習結果を用いた自己連想の例

## 5. おわりに

本稿では、汎化性能を向上させるための新たな学習アーキテクチャとして、通常のニューラルネット識別器の中間層に (1) 入力自己連想を同時に学習させ、さらに (2) 学習過程で中間層各素子に独立なノイズを付加する、という方法を提案し、提案手法の有効性を「照明変動下の顔画像識別問題」で示した。

今後の課題として、提案手法の有効性を他の識別問題でも確認すること、および、汎化性能向上の機構解明すなわち提案手法でなぜ/どんな仕組みで汎化性能が向上するのかの解明に取り組む予定である。

## 文 献

[1] 高橋隆史 栗田多喜夫 池田幸史 “画像の例外部分を補正する前処理機構をもったニューラルネット識別器”, 電子情報通信学会論文誌 D-II vol.J87-D-II no.5 pp.1162-1169, 2004  
 [2] 栗田多喜夫, 麻生英樹, 梅山伸二, 赤穂昭太郎, 細美章隆, “多層

パーセプトロンの学習における中間層に付加したノイズの影響とネットワークの構造化,” 電子情報通信学会論文誌, Vol.J79-D-II, No.2, pp.257-266, 1996 (in Japanese).

[3] Bishop, C. M., “Training with noise is equivalent to Tikhonov regularization”, Neural Computation 7(1), pp.108-116, 1995.  
 [4] Akaho, S., “Regularization Learning of Neural Networks for Generalization”, Algorithmic Learning Theory (ALT’92: Lecture Notes in Artificial Intelligence, Vol.743 Springer-Verlag), 1992.  
 [5] A.F. Murray, P.J. Edwards, “Synaptic Weight Noise During MLP Training : Fault Tolerance and Training Improvements”, IEEE Trans. Neural Networks, 4, No 4, pp. 722-725, 1993  
 [6] J.B. Tenenbaum, T.L. Griffiths “Generalization, similarity, and Bayesian inference”, Behavioral and Brain Sciences, 24 pp. 629-641. 2001.  
 [7] 岡部孝弘, 佐藤洋一, “照明変化を伴う顔認識”, 情報処理学会研究報告, CVIM 2003-139-10, pp.69-76, July 2003.  
 [8] Adini, Y and Moses, Y and Ullman, S, “Face Recognition: the Problem of compensating for Changes in Illumination Direction”, Technical Report CS93-21, Mathematics & Computer Science, Weizmann Institute Of Science, 1993  
 [9] Belhumeur, P.N., Kriegman D.J., ”What Is the Set of Images of an Object Under All Possible Illumination Conditions?”, Int. Journal of Computer Vision, 28(3),PP. 245-60, 1998  
 [10] M. Turk, A. Pentland, “Eigenfaces for Recognition”, Journal of Cognitive Neuroscience, 3(1), pp71-86, 1991  
 [11] Belhumeur, P.N., Hespanha, J., Kriegman D.J., “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, IEEE Trans. PAMI, Special Issue on Face Recognition, 19(7),pp. 711-20 1997  
 [12] M. Savvides, B.V.K.V. Kumar, P.K. Khosla, “ ”Corefaces” - Robust Shift Invariant PCA Based Correlation Filter for Illumination Tolerant Face Recognition”, CVPR’04 pp.834-841, 2004  
 [13] Georghiades, A.S. and Belhumeur, P.N. and Kriegman, D.J.”, “From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose”, IEEE Trans. Pattern Anal. Mach. Intelligence”, 23(6), pp.643-660, 2001