

自己組織化マップ (SOM) によるモザイク顔画像認識

山森 一人¹⁾・平山 政宏²⁾・吉原 郁夫³⁾Mosaic Face Image Recognition by
Kohonen's Self Organizing MapsKunihito Yamamori¹⁾, Masahiro Hirayama²⁾, Ikuo Yoshihara³⁾

Abstract

In recent years, identification of persons becomes very important for security. Password has been generally used for this identification, however, persons sometimes forget their password, and sometimes it is stolen. Therefore a new technology is required for robust identification. In this paper, we use mosaic face images to recognize persons by self organizing map (SOM) proposed by Kohonen. In addition we evaluate the SOM based recognition by using of Euclid distance and city block distance. The SOM with the Euclid distance succeeded complete recognition, and the one with the city block distance achieved 96.7% recognition ratio.

Key Word:

Neural network, Self-organizing-maps, Face image recognition, Mosaic

1. はじめに

現在の情報化社会において、個人認証は銀行のキャッシュカードや建物の出入り、コンピュータへのログインなど多方面で重要性を高めている。しかし、従来から使われているパスワードによる認証では、パスワードを忘れて盗まれたりするといった問題がある。この問題を解決する方法として、指紋、虹彩、音声、DNA、顔画像など人間の生体情報（バイオメトリクス）を用いた個人認証が研究されている。指紋などは個人を示す情報として犯罪捜査などにも用いられ、信頼できる認

証法である。しかし、指紋を採取される側にとっては精神的にストレスを感じやすい。なぜなら、普段目にしない機械や道具を使用するうえ、特に指紋などは犯罪捜査のイメージが強いためである。一方、顔画像は特別な道具を使わずストレスも与えにくいといった長所をもっており、今後ますます活用されると考えられる。

本研究ではニューラルネットワークの一種である自己組織化マップ(SOM)を用いて顔画像認識を実現することを目的とする。SOMでは、学習によってトポロジカルマップを生成できる。トポロジカルマップとは、類似した入力に対し近傍位置にあるニューロンが大きい出力を出すよう、入力情報の位相をノード間の位置関係に対応させたも

1) 情報システム工学科助教授

2) 情報システム工学科学部生 (現職, 船井電機株式会社)

3) 情報システム工学科教授

のである。これにより、撮影時のノイズや位置ずれがあった場合でも近傍ノードが活性化するので、よりロバストな認識が可能になる。さらに、顔画像にはモザイク化処理を施すことで、顔の表情の微妙な違いといったわずかな誤差についても対応可能とする。

本論文の構成は以下の通りである。2章でSOMの学習アルゴリズムについて説明し、3章では本研究で用いた顔画像認識の手法を示す。4章では実験について述べ結果について考察を行う。5章はまとめである。

2. ニューラルネットワーク (NN)

2.1 自己組織化マップ

ニューラルネットワークは生物の神経回路をもとに作られた情報処理モデルである。本研究では、コホネン[1]が提案した自己組織化マップ(Self Organizing Maps, 以下SOM)を用いる。パーセプトロンや誤差逆伝搬法が「教師あり学習」であるのに対し、SOMは競合学習型ニューラルネットワークであり、その学習は入力信号の性質のみに基づく「教師なし学習」である[2]。SOMはトポロジカルマッピングとよばれる性質を持ち、類似した入力、例えば同じ人物の違う顔画像に対し、学習によって形成される2次元のマップ上において近傍に位置するノードが活性化する。これにより、活性化したノードの位置情報を用いて個人認証が可能となる[3][4]。

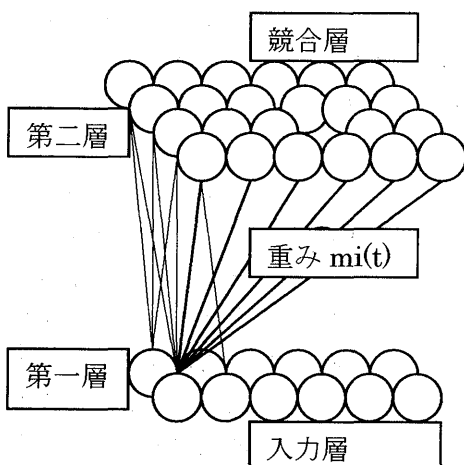


図1 SOMの構成

2.2 SOMの学習アルゴリズム

図1に示すようにSOMは2層構造である。第1層は入力層であり、 n 個の入力ノード各々が入力1つを受け持つ。第2層は競合層と呼ばれ、ノードが格子状に並べられている。競合層のノードは入力層のノードと完全結合している。時刻 t で競合層のノード i が持つ重みは $m_i(t)$ で表され、これは入力層のノード数と同じ n 個の要素を持つベクトルである。また、学習の基準となる測度には入力 $x(t)$ と $m_i(t)$ 間のユークリッド距離が一般に用いられる。

SOMの学習は(1)式で表される。

$$m_i(t+1) = m_i(t) + hc_i \{x(t) - m_i(t)\} \quad (1)$$

(1)式で hc_i は(2)式で定義される係数である。

$$\begin{cases} hc_i = \alpha(t) & (i \in N_c(t)) \\ hc_i = 0 & (i \notin N_c(t)) \end{cases} \quad (2)$$

ここで、 c は入力 $x(t)$ に最も類似した重みを持つ競合層ノードの番号であり、(3)式で求められる。

$$c = \underset{i}{\operatorname{argmin}} |x(t) - m_i| \quad (3)$$

また $\alpha(t)$ 、 $N_c(t)$ は(4)式、(5)式でそれぞれ定義される。

$$\alpha(t) = \alpha(0) \left(1 - \frac{t}{T}\right) \quad (4)$$

$$N_c(t) = N_c(0) \left(1 - \frac{t}{T}\right) \quad (5)$$

(4)式、(5)式において、 T は最大学習回数を表し、 $\alpha(0)$ は初期学習率係数と呼ばれ通常0.1~0.5の値をとる。また、競合層上で重み更新の範囲を表す $N_c(t)$ は近傍範囲と呼ばれる。これは(3)式により求められた勝者ノード c を中心とした、競合層上の半径 $N_c(t)$ の矩形領域であり、初期値 $N_c(0)$ は一般に競合層の半径よりも大きくとる[1]。

2.3 SOMの学習の流れ

SOMの学習の流れは以下の通りである。学習用入力 $x(t)$ が提示されると、 $x(t)$ と競合層のノードの重み m_i との間でユークリッド距離をとり、

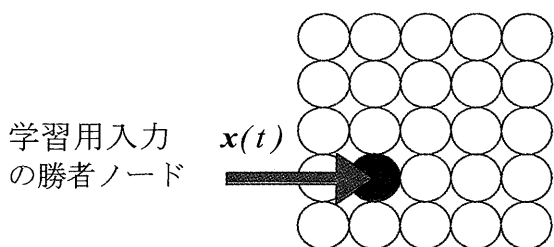


図2 勝者ノードの決定

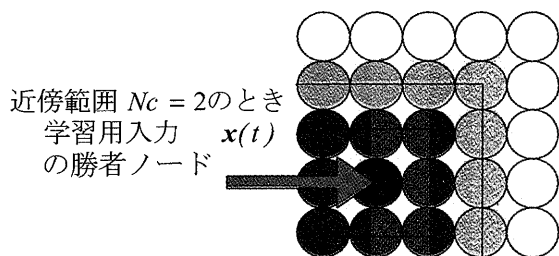


図3 近傍範囲の学習

図2に示すように最もその距離の短いノードを勝者ノード c とする. 次に, 図3のように c の近傍範囲を(5)式によって求め, 近傍範囲中の全ノードの重みを(1)式により $x(t)$ に近づく方向へと更新する. また, 図3では濃い色ほど大きく重みを更新したことを表し, 薄い色ほど更新量が小さくなっている事を表している.

3 顔画像処理 (モザイク化処理)

3.1 顔画像の撮影

本研究に用いる顔画像は, 同じ場所, 同じ距離, 同じ時間に撮影し, 明るさや撮影距離に大きな違いがでないよう留意した. 顔画像は一人につき学習用として4枚, 認識用として1枚の計5枚を撮影した.

3.2 顔画像の切り出し

本研究では, 256pixel×256pixelのサイズで顔領域の切り出しを行った. 切り出し自体は手作業で行い, 顔の中心部を切り出した. 切り出した画像はpgm形式(グレースケール256階調)に変換した.

3.3 顔画像のモザイク化処理

モザイク化処理は, 256階調に変換した顔画像を縦横 n 等分に分割し, 各部分領域内の画素の平均輝度値を各部分領域の輝度値とすることで行った. モザイク化の例を図4に示す. 本研究では過去の実験結果より $n=8$ としている[5].

SOMの入力には実数値を用いることができるので, 本研究ではモザイク化により得られた輝度値をそのまま入力値として使用する. また, 図4に基づいた実際のSOMへの入力ベクトルを表1に示す.

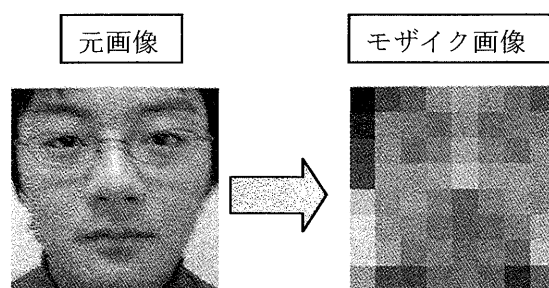


図4 モザイク化処理

表1 モザイク化によって得られた入力データ

23	74	97	151	173	146	117	92
45	134	124	108	148	110	126	127
69	145	107	133	176	135	116	150
72	160	154	168	200	162	162	152
204	148	162	136	102	132	157	143
231	157	141	118	102	109	132	177
226	173	119	137	110	125	112	201
160	81	76	110	117	105	59	107

表2 学習用と認識用の入力パターン

	学習画像				認識画像
	No.1	No.2	No.3	No.4	
パターン1	No.1	No.2	No.3	No.4	No.5
パターン2	No.2	No.3	No.4	No.5	No.1
パターン3	No.3	No.4	No.5	No.1	No.2
パターン4	No.4	No.5	No.1	No.2	No.3
パターン5	No.5	No.1	No.2	No.3	No.4

4 SOMによる顔画像認識実験

4.1 学習条件

実験にあたり、一人につき5枚ずつ12人分の計60枚の画像を用意した。1名分の画像5枚の各々をNo.1~No.5とする。5枚のうち4枚をSOMでの学習に用い、残りの1枚を認識用として生成したマップ上に配置し認識結果を調べた。学習に用いた画像による偏りを防ぐため、学習用画像、認識用画像は表2に示すような5パターンを用意した。

4.2 認識方法

本研究では認識方法として以下に述べる2つの方法を比較した。

4.2.1 ユークリッド距離による認識

本方法では、まず、認識させたい未知入力と競合層のノードの重みとのユークリッド距離をとり、最も距離の短いノードを勝者ノードとする。次に学習により画像と対応づけられたノードのうち、勝者ノードとの重みのユークリッド距離が最も短いノードが対応する人物を認識結果とする。

4.2.2 シティブロック距離による認識

シティブロック距離は、2次元格子上で2点の座標を $(x_1, y_1), (x_2, y_2)$ とすると(7)式で表される。

$$|x_1 - x_2| + |y_1 - y_2| \tag{7}$$

ユークリッド距離を用いたときと同様に、はじめに未知入力と競合層のノードの重みとのユークリッド距離をとり、最短距離となるノードを勝者ノードとする。次に、勝者ノードから各人物の4枚の画像に対応するノードまでのシティブロック距離を求め、その平均をとる。求めた平均距離が最短となる人物を未知入力の認識結果とする。

4.3 初期パラメータの決定

本研究では、初期近傍範囲の値は競合層の半径に+2を加えた値、つまり $N_c(0)=7$ とした[1]。 $\alpha(0)$ の値は、シティブロック距離を用いるときの汎化能力低下による誤認識を防ぐため、同一人物の4枚の画像それぞれが別のノードに割り当てら

れるよう設定することが望ましい。そのため、 $\alpha(0)$ を0.1~0.5まで0.1刻みで値を変化させ、各人毎に割り当てられたノード数を調べた。その結果を表3に示す。表3より $\alpha=0.5$ が最適であることが分かるが、同時に4枚の画像全てが同一ノードに割り当てられたケースも多かった。そこで、今回は次点である $\alpha=0.3$ を全ての実験で使うことにした。また、競合層サイズは 10×10 とし、顔画像12人分の学習用画像48枚を表4に示す学習パラメータで学習させた。

4.4 実験結果

図5に、パターン1で10000回の学習を行って得られたマップの1つを示す。マップ上のアルファベットA~Lはそれぞれ12人の顔画像に対応しており、一人当たり4枚の計48枚が配置されている。

表3 $\alpha(0)$ による競合ノードの重複の違い

	各人に割り当てられた競合ノード数			
	4個	3個	2個	1個
$\alpha=0.1$	1	9	1	1
$\alpha=0.2$	2	8	0	2
$\alpha=0.3$	3	7	1	1
$\alpha=0.4$	1	9	1	1
$\alpha=0.5$	4	5	0	3

表4 学習に用いたパラメータ

	$\alpha(0)$	$N_c(0)$	競合層サイズ
パラメータ値	0.3	7	10×10

H		G	C		G	G		E	I
H	H		B		K		F		F
			B						
L	L		B		K		E		F
					K		F		
L		B	B		K				J
					K		J		
A	A	D	D						
				D	D		J		J

図5 認識により得られたマップ

同一人物の異なる画像が同じノードに対応することがあるので、48枚の画像が競合層のノードに1対1で対応しているわけではない。また、図5では学習画像での勝者ノードをローマン体で、認識画像での勝者ノードを下線付きで、それらが重複した場所をイタリック体で示している。

4.4.1 ユークリッド距離を用いた認識

パターン1でユークリッド距離により認識を行った結果を表5に示す。表5中の一位候補とは、未知画像での勝者ノードの重みに最も近い重みを持つノードに対応する人物であり、一位候補との距離はそのユークリッド距離を示す。表5より、12人(A~L)全てのモザイク顔画像を正確に認識できていることが分かる。また、一位候補と二位候補の間では距離に大きな差があり、ロバストな認識が実現できていることが分かる。

表5 ユークリッド距離を用いた認識

画像	A	B	C	D	E	F
1位候補	A	B	C	D	E	F
1位候補との距離	58.8	53.7	43.5	53.6	24.1	49.1
2位候補	D	F	G	H	G	K
2位候補との距離	307.8	246.1	176.4	217.3	131.1	188.2
3位候補	L	G	K	L	H	G
3位候補との距離	323.3	252.4	180.7	225.9	179.5	203.1
画像	G	H	I	J	K	L
1位候補	G	H	I	J	K	L
1位候補との距離	21.1	30.2	49.6	78.5	41.8	47.9
2位候補	E	E	G	F	C	D
2位候補との距離	141.1	167.4	188.5	414.8	184.3	168.1
3位候補	C	J	E	K	G	G
3位候補との距離	174.0	168.1	194.3	427.1	190.1	208.1

4.4.2 シティブロック距離を用いた認識

シティブロック距離を用いて認識を行った結果を表6に示す。表の見方はユークリッド距離の場合と同様であり、パターン1のときは全てのモザイク顔画像を正確に認識することができた。この

表6 シティブロック距離を用いた認識結果

画像	A	B	C	D	E	F
1位候補	A	B	C	D	E	F
1位候補との距離	0.5	2.25	0.5	2.0	0.0	2.25
2位候補	D	C	H	J,K	I	K
2位候補との距離	3.0	3.5	3.25	4.25	1.0	3.75
3位候補	L	K,H	G	J,K	G	J
3候補との距離	3.75	4.25	3.75	4.25	3.25	4.75
画像	G	H	I	J	K	L
1位候補	G	H	I	J	K	L
1位候補との距離	0.25	1.75	0.0	2.25	1.75	1.25
2位候補	E	C	E	K	B	A
2位候補との距離	2.0	2.5	1.0	4.25	3.75	2.5
3位候補	I	L	G	F	F	B
3候補との距離	3.0	5.25	3.25	4.75	4.75	4.25

ことから、学習によりトポロジカルマップが正しく生成されていることが分かる。

4.4.3 認識手法の比較

表7に各入力パターンでのユークリッド距離による認識率とシティブロック距離による認識率、および各々の平均認識率を示す。

表7より、シティブロック距離を用いた認識での平均認識率は96.7%であり、パターン4とパターン5で各々1人について誤認識が生じた。

パターン4では、一位候補が2人になり正しい認識ができなかった。これは、人物A, Bに対応するノードa, bが図6のように配置され、Aの未知画像A'に対応する勝者ノードがa'に配置された場合に相当する。aのノード群、bのノード群ともa'からの平均シティブロック距離が等しいため、A'がAに属するのかBに属するのか判断できず誤認識となった。

パターン5では、正解とは別の人物が一位候補となり、正しい人物が二位候補となっていた。これは、人物A,Bに対応するノードa,bが図7のように配置され、Bの未知画像に対応する勝者ノードがb'に配置された場合に相当する。図7より、

表7 各手法での認識率

入力 パターン	ユークリッド距離		シティブロック距離	
	正解人物数	認識率 (%)	正解人物数	認識率 (%)
1	12	100	12	100
2	12	100	12	100
3	12	100	12	100
4	12	100	11	91.7
5	12	100	11	91.7
平均	12	100	11.6	96.7

Aの画像4枚が対応付けられたノードaへの(平均)シティブロック距離が2である一方、Bの画像2枚が対応付けられたノードb(2個)までの平均シティブロック距離が3となるため誤認識が生じる結果となった。

これらのことから、シティブロック距離を用いた認識では異なるグループの画像は競合層上で十分な間隔を持って配置される必要があること、同一グループに属する画像は競合層ノードに重複することなく隣接して配置する必要があることが明らかとなった。

5. おわりに

現代のような情報化社会のなかで重要度を増す個人認識を安全に行う方法として、個人の生体情報(バイオメトリクス)を用いて認識を行う方法が注目されている。本研究ではデータの取得の容易さなどの長所を持つ顔画像に着目し、ノイズに対する耐性を強めるためにモザイク化を行ったうえでSOMによる認識を試みた。認識においてはユークリッド距離を用いる場合とシティブロック距離を用いる場合の2つを比較した。ユークリッド距離を用いた場合、12名の顔画像すべてを正しく認識することができたが、シティブロック距離を用いた場合一部で誤認識を生じたため、その原因について考察を行った。

今後の課題として、より多人数の認識を行うこ

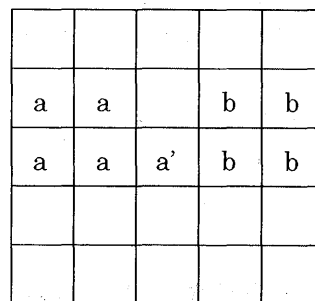


図6 対称に配置されたノードによる誤認識の例

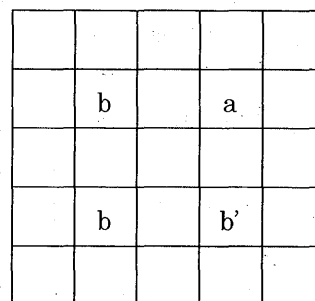


図7 距離平均化による誤認識の例

とや撮影時の位置ずれなどへの耐性について調査すること、シティブロック距離を用いた時の認識精度向上が挙げられる。

謝辞

本文に掲載する顔画像を提供してくれた藤田祐司氏に感謝します。

参考文献

- [1]T.コホネン, (徳高平蔵, 岸田悟, 藤村喜久朗 訳) “自己組織化マップ”, シュプリンガー・フェアラーク東京(1996)
- [2]古富康成 “ニューラルネットワーク”, 朝倉書店(2002)
- [3]徳高平蔵, 藤村喜久朗, 山川烈 監修 “自己組織化マップ応用事例集”, 海分堂(2002)
- [4]徳高平蔵, 岸田悟, 藤村喜久朗 “自己組織化マップの応用”, 海分堂(1999)
- [5]野川玲雄 “モザイクを利用した顔画像認識” 平成14年宮崎大学工学部 卒業論文(2003)