

顔の大きさに依存しないニューラルネットワークを用いた顔領域の切り出し*

4M-4

三浦 寛典

有木 康雄

龍谷大学 理工学部

1 はじめに

画像中から人の顔領域を切り出す処理は、個人識別を行なうセキュリティシステムにとって必要であるとともに、画像の情景を理解する上でも必要である。顔領域の切り出し及び認識に関する研究では、顔の濃淡画像をモザイク化する手法が提案されている [1] [2]。手法 [1] では、複数人の顔領域を切り出すことはしておらず、また、白髪及び髪の毛の薄い人には失敗してしまうという問題がある。

本研究では、1枚の濃淡画像をモザイク化して直流成分で画像を表現し、ニューラルネットワークを用いてシーン画像から複数人の顔領域を切り出す手法を提案している。また、モザイクでは切り出しが難しいと思われる白髪の顔領域を抽出するために、画像の直流成分のみではなく、交流成分と合わせて画像を表現し、顔領域を切り出す手法も試みたのでその結果を報告する。

以下、学習用データの作成、ニューラルネットワークの学習方法、顔領域の切り出し方法と結果について述べる。

2 学習用データの作成

顔領域をニューラルネットワークによって精度よく切り出すために、ニューラルネットワークには顔領域のデータ（顔データ）だけでなく、顔でないデータ（非顔データ）を学習させている。顔データは、学生アルバムの中から男27人、女23人を選び、モザイク化及びDCTを行なって作成した。画像の大きさは 90×72 画素であり、モザイク化及びDCTのために 10×8 個のブロックに分割している。モザイク化では、ブロック内で濃淡値の平均（直流成分）を求める。DCTではブロック内で直流成分と縦・横の最も低い周波数成分、合わせて3成分を計算する。

顔は髪の毛を含む正面像であり、ランダムに男女ともメガネ着用者を含めている。

非顔データは、自然画像中から顔以外の領域を切り出したものであり、顔データと同様に、モザイク化及びDCTを行なって作成した。画像サイズは顔データ

と同じである。

また、照明条件の差を軽減するために、ヒストグラムの平坦化処理を施し、画素の濃淡を画像全体に拡大して、濃淡補正を両データに対して行なっている。

3 ニューラルネットワークの学習方法

本研究で用いたニューラルネットワークの構造は図1に示すように3層である。

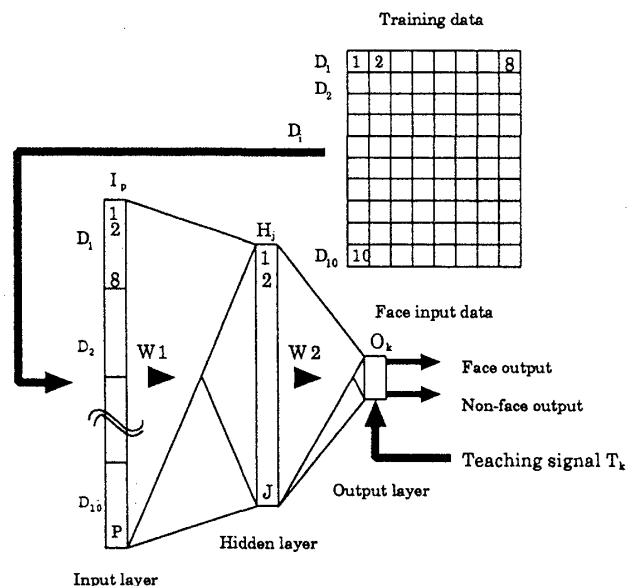


図1: ニューラルネットワークの構成図

学習用データは、 9×9 画素の領域を1ブロックとして、 10×8 個のブロックに分割されている。モザイク化を行なった学習データに対して、横方向の8ブロックをまとめて D_i とし、上から順に並べて、1次元配列を作る。この1次元配列をニューラルネットワークの入力層に入力している。従って入力層のノード数は80である。

出力層は顔出力と非顔出力の2つのノードがある。この理由は、顔と背景を含んだシーン画像より顔領域を精度良く切り出すために、顔の特徴のみを学習するのではなく、背景と顔領域を識別することを目的としているためである。すなわち顔らしさ、背景らしさの識別機能の一つのネットワークで学習させている。中間層のノード数は102に設定している。ニューラルネットワークの学習方法は誤差逆伝搬法である。

*Size Independent Extraction of Facial Region by Neural Networks.

Hironori Miura and Yasuo Ariki (Ryukoku University)

DCTを行なった学習データでは、直流成分に対しては図1と同じニューラルネットワークを用いて学習を行なう。交流成分に対しては、入力層のノードの数が、 $10 \times 8 \times 2$ のニューラルネットワークを用いて学習を行なう。

4 シーン画像からの顔領域の切り出し

シーン画像より顔領域を探索する方法は、まず図2に表されるように、学習データの入力ブロック数 10×8 と同じ大きさの領域を最上層部とする多階層ピラミッドを形成する。最上層部のブロックは 1×1 画素で構成されており、ブロックの大きさはピラミッドが下方に行くに従い、 2×2 、 $3 \times 3 \cdots$ と大きくなっていく。つまり、ピラミッドの a 層では $a \times a$ 画素の大きさのブロック内で濃度値の平均を求めてモザイク化することにより、 $10 \times 8 \times a^2$ の大きさの領域についても探索できるようになっている。この方法により、入力層のノード数は80と固定であるが、顔の大きさに依存しないで、顔領域を切り出すことができる。このピラミッドを画像全体に適用し、ニューラルネットワークの出力値を計算する。顔ノードの出力値が高ければ、その領域を顔と判定し、顔領域の切り出しを行なう。

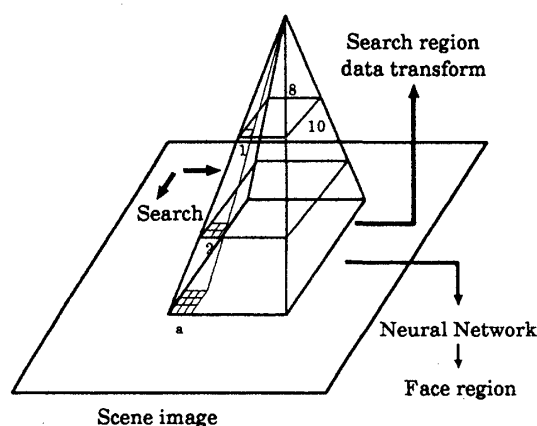


図2: 多階層ピラミッドサーチ

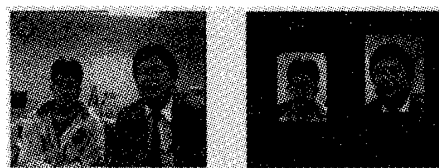
5 実験評価

モザイクによる顔領域の切り出し結果の1例を図3に、DCTによる切り出し結果の例を図4に示す。

また、未学習データ50サンプルに対する定量評価として、モザイクによる切り出し実験の結果を表1に示す。表1で検出率とは、真に存在する顔領域のうちで正しく切り出した顔領域の割合であり、検出もれとは、切り出せなかった顔領域の割合である。また、誤検出とは、顔として切り出した顔領域の中で、顔でない領域の割合である。

DCTの切り出しについては、ブロックサイズ a が

学習したサイズよりも小さくなると、周波数解像度の違いから失敗する傾向がある。



(a) 原画像 (b) 切り出し結果

図3: モザイクによる顔領域切り出し



(a) 原画像 (b) 切り出し結果

図4: DCTによる顔領域切り出し

表1: 検出率

検出率	検出もれ	誤検出
74%	26%	41.3%

6 むすび

モザイク化及びDCTを行なった画像に対して、ニューラルネットワークを用いて顔データと非顔データを学習し、シーン画像から顔の存在場所や大きさに依存しないで顔領域を切り出す方法を提案した。モザイクによる切り出しでは、眼鏡の有無や、性別に依存しないことが確かめられた。また、モザイクで失敗するような場合には、ブロック内の交流成分も含めて学習させることにより、切り出し精度が向上できることを確認した。しかし、ブロックサイズと周波数解像度の違いが問題となっており、今後この問題を解決する予定である。

参考文献

- [1] 小林 信: "シーンの中の顔の探索と認識", 信学技報, PRU91-104, (1991).
- [2] 小林 信: "モザイクとニューラルネットワークを用いた顔画像の認識", 信学論, J76-D-II No.6, pp.1132-1139, (1993).