氏名（本籍）	糸井　清晃　　（埼玉県）
学位の種類	博士（工学）
学位記番号	乙第87号
学位授与の日付	令和元年9月12日
学位授与の要件	学位規則第4条第2項該当
学位論文題目	形態を指向した非線形画像処理に関する研究
論文審査委員	（主査）教　授　　中静　　真
	（副査）教　授　　久保田　稔
	教　授　　菅原　真司
	教　授　　枚田　明彦
	教　授　　宮田　高道

学位論文の要旨

形態を指向した非線形画像処理に関する研究

本論文では,形状の処理を目的とした画像処理に関して幾つかの手法を提案する.画像処理は,処理結果として出力される情報の種類によって，大きく二つに分けられる．画像が出力される処理を（狭義の）画像処理と呼び，一方，画像を表現するものの，画像形式ではない情報が出力される処理を画像解析（計測・認識）と呼ぶ．前者の一部は，画像を時間的または空間的に変化する色情報のディジタル信号と捉え，フィルタ処理として，もしくは，画素値の並びを行列と捉えれば行列演算として，処理を実現することが可能である．一方，被写体がなにであるかが意味を持つような処理は困難な場合が多い．このような場合には，被写体の形状を考慮した処理が必要になる．例えば，文書画像の処理であれば，文書を構成する主たる要素である文字の形状が重要であり，顔画像処理であれば，顔を構成する要素の形状が重要である．このような処理を実現する方法には，画像が持つ形状の特徴を人間の知識に基づいて解釈し，対象となる画像または所望する画像処理に特化したアルゴリズムを構築して処理する方法と，機械学習によって，コンピュータを複数の事例を用いて学習させ，未知の入力を処理する方法がある．

本研究では，文書画像と顔画像に関する形状を考慮した画像処理について検討する．文書画像に関しては，紙媒体の文書を電子化する際に有用となる処理について検討している．文書を電子化しテキストデータに変換する処理は，文書を保存する際の省スペース化及び情報検索の効率化の面で重要である．本論文では，文書を文字認識してテキストデータ化する際に障害となる，

見出し文字列の背景に付けられた「地紋」と呼ばれる飾りを除去する方法と，印刷文書に書き

込まれた注釈などから有益な情報を得るために，手書き文字を抽出する方法を提案している．前者では，多数の見出し文字列を調査して形状の特徴によって 3 種類に分類し，分類されたそれぞれの地紋を除去するのに特化したアルゴリズムを検討して，個々のアルゴリズムを組み合わせることで全体のアルゴリズムを構成している．後者では，形状を除去することができるモフォロジカルフィルタの一つであるクロージングフィルタ用いるが，その性能を向上させるために，フィルタを構成するダイレーションを Maxout 関数によって拡張し，Maxout フィルタネットワークを構成する．このフィルタネットワークを 3 層のニューラルネットワークとみなし，その係数を多数の事例によって学習させることによって，手書き文字だけを抽出できるようにしている．学習方法は，従来の「確率的勾配降下法」「ミニバッチ学習」に加えて，「ミニバッチ＋最大値学習」を提案している．最大値学習とは，ニューラルネットワークの入力と出力を比較し，大きい方を改めて出力とする学習方法である．以上の三つの方法を用いて学習を行い，その性能を比較している．学習のための事例となるデータは，手書きの注釈が書き込まれた印刷文書画像を入力データ，それに対応する手書きの注釈のみの画像を「理想手書き文字」と呼び，教師データとしている．

顔画像処理では，テレビ電話やテレビ会議システムなどの通信において，話者の表情を分析することによって低ビットレート化を実現する知的符号化を提案している．提案手法では，送信側のカメラによって動画として撮影された話者の表情を分析した結果を極少ない数の数値として送信し，受信側で受信した表情の分析結果の数値から送信話者の表情に似た表情を合成してディスプレイに表示する．表情の分析では，話者の特徴的な表情を選んで基本表情とし，これらを階層型ニューラルネットワークで学習させておく．このニューラルネットワークによって，カメラで撮影された話者の任意の表情と基本表情との類似度を計算する．表情の合成は，表情の分析結果である類似度を基に計算される割合によって，基本表情をモーフィングを用いて合成することによって実現する．

提案した手法に関して実験を行い，次のように評価し，その有効性を確認している．

地紋の除去では，処理結果である地紋が除去された見出し文字列を目視と文字認識率によって評価している．目視による評価では，処理結果を 4 段階で評価し，概ね良好に除去できることを確認している．文字認識率による評価では，目視による評価それぞれに属する処理結果ごとに認識実験を行い，認識率が目視評価と同じ傾向となることを確認している．また，目視により良好と評価されたものの認識率は，元々地紋のない見出し文字の認識率とほぼ同等であることも確認している．

手書き文字の抽出では, 目視と SNR によって評価している. SNR の平均は, 確率的勾配降下法, ミニバッチ学習，ミニバッチ＋最大値学習の順に大きくなっており，提案した最大値学習の効果を確認している．また，目視による評価では，理想手書き文字と比較して遜色ない抽出結果になっていることを確認している．更に，従来の印刷文書からの手書き文字抽出手法では困難であった，印刷文字と手書き文字が重なっている場合の処理も可能となったことを確認している．知的符号化では，受信側で出力される合成表情が，概ね送信側のカメラで撮影した話者の表情の変化

に追従して変化していることを確認している．また，提案手法による顔画像の送信に必要なデータ構造（ビット数）に関する検討も行い，画像のまま送信する場合に比べて，極超低ビットレートを実現できることを確認している．

審 査 結 果 の 要 旨

本論文は，画像処理の中でも，特に形状，および複数の形状から構成される形態に意味を持つ画像である文字画像および顔画像に対する処理について，「形態を指向した非線形画像処理」として研究をまとめたものである．

第 1 章では，論文全体を通しての背景と目的について述べている．文書画像処理は，紙メディア上に記されたコンピュータには直接読み取れない情報を，コンピュータ上で取り扱うための研究分野であり，総じて画像として入力された文書から，コンテンツを抽出することが主な目的である．文字情報を抽出し，文字認識により文章をデータとして保存するためには，文章画像中にある不要な要素を除去する必要がある．また，メモとして手書き文字などが重畳した文書画像では，手書き文字と活字を分離して記録，アーカイブへ保存する必要があり，形状に着目した分離法が必要となる．

また，人と人の間でコミュニケーションを図る上で，顔の画像情報は相手の個人の特徴だけでなく，その表情によって心理状態など，声だけでは伝わりにくい様々な情報を伝えるという役割を担っている．遠隔地の人間とこのようなコミュニケーションをとるために提案されたテレビ電話・テレビ会議システムにおいては，映像が伝えるべき重要な情報は，話者の表情である．映像の圧縮技術が進歩した現在においても，映像情報の伝達はコストのかかる処理であるため，顔のパーツそれぞれの形状を組み合わせた形態である表情を，少ないデータ量で伝達する必要性がある．第 1 章では，社会的な要求に基づき，これら形状と形態に基づく画像処理の必要性と問題設定について述べ，研究遂行の上での問題設定を示している．

第 2 章では，画像として保存されている新聞の画像を対象として，文書検索のキーワードとなり得る単語を有する見出しの文字認識の際に障害となる「地紋」と呼ばれる背景の飾りを除去する方法を提案している．既存の方法では，除去できない地紋を，詳細に分類し，その分類に応じた処理を検討することで, 多種の字紋を除去することができた. この分類と除去アルゴリズムを，目視とコンピュータによる文字認識精度から評価を行っている．その結果，提案法は，見出し文字の認識精度を向上させ，100%に近い認識率が達成できることを示した．

第 3 章では，文章，特にメモ等の手書き文字を含んだ文章をアーカイブに保存するために，印刷文章からの手書き文字の抽出について提案した．第 2 章で解決した問題で除去すべき対象であった地紋と文字では，大きく形態が異なるため，手動による分類により地紋を区別し，アルゴリズムを構築することができた．それに対して，手書き文字の抽出の問題は，形状が類似した印刷文字を除去する問題となり，第 2 章で採用したアプローチでは，解決できない問題である．そこ

で，既存の画像処理のための非線形フィルタであり，形状選択性を有するモフォロジカルフィルタに着目し，これをニューラルネットワークの活性化関数である Maxout 関数を用いて訓練可能な 3 層ネットワークに拡張することで，学習に基づく処理を提案した．

文字抽出の問題の性質から，既存の 3 層ネットワークに新たなパスを加えた新しいネットワーク構造を提案し，学習のためのバッチ処理の方法を検討した．その結果，少ない学習データで有意な処理結果を得ることができた．さらに，実験により多種の手書き文字に適用することが可能で，既存の手法では抽出が困難であった重なりのある文字の分離に対しても良好な結果が得られることを示した．

第 4 章では，顔画像からの表情分析と，表情データを伝送することで超低ビットレート顔画像符号化を提案した．顔画像の送信側の表情分析では，話者の特徴的な表情を選んで基本表情とし，これらをニューラルネットワークに学習させる．

このニューラルネットワークによって，カメラで撮影された話者の任意の表情と基本表情との類似度を計算し，その結果を送信する．受信側では，送信話者の表情の分析結果である類似度を基に，モーフィングを用いて基本表情を合成する．実験によって，受信側で出力される合成表情が，送信側のカメラで撮影した話者の表情の変化に追従して変化し，円滑な顔の表情変化を伝送できることを示した．また，提案手法による顔画像の送信に必要なデータ構造（ビット数）に関する検討も行い，画像のまま送信する場合に比べて，ニューラルネットワークから抽出された表情データを伝送することで，極超低ビットレートの顔画像伝送を実現できることを確認した．

第 5 章では，文字画像および顔画像を対象とした研究成果の総括を行い，ニューラルネットワークを中心とした非線形画像処理に関する有益な知見を整理し，残された課題を示した．

これらの研究を通じて，形態を指向した画像処理に関する包括的な知見を得ることができた．これらは，特に文字画像からの文章のアーカイブ，通信における表情伝送の分野において，有益な事項といえる．

以上のことから，本研究は画像処理分野，メディア処理分野について新たな重要な知見を与えるものとして価値ある長年の集積であると認め，学位申請者である糸井清晃は，博士(工学)の学位を得る資格があると認める．