

平成 26 年度 修士論文

デジタルマンガ画像における顔検出に関する研究
Research on Face Detection of Digitized Comic Images

指導教授 渡辺 裕 教授

2014 年 1 月

早稲田大学大学院 国際情報通信研究科

国際情報通信専攻

4213A154-5

陳 明

Ming CHEN

目次

第 1 章 序論	1
1.1 研究の背景	1
1.2 研究の目的	2
1.3 本論文の構成.....	3
第 2 章 電子コミックの現状	4
2.1 はじめに.....	4
2.2 電子コミックの発展史	4
2.3 電子書籍マーケットの現状	5
2.3.1 電子コミック市場	5
2.3.2 紙からデジタルへの転換.....	6
2.4 おわりに	8
第 3 章 画像局所特徴量	9
3.1 はじめに	9
3.2 「パターン認識」による顔検出の基本的な流れ.....	9
3.3 HAAR-LIKE 特徴量	10
3.3.1 局所的な明暗差の集合	10
3.3.2 HAAR-LIKE 特徴パターンの分類.....	11
3.3.3 HAAR-LIKE 特徴の計算.....	11
3.3.4 HAAR-LIKE 特徴の積分画像	12
3.3 SIFT 特徴量.....	12
3.3.1 SIFT の長所	13
3.3.2 SIFT 特徴で処理できる問題[15]	13
3.3.3 SIFT 特徴とコミック画像の親和性	13
3.4 HOG 特徴量	14

3.4.1	HOG 特徴の算出アルゴリズム.....	14
3.4.2	HOG 特徴の次元数算出	16
3.5	LBP 特徴.....	16
3.5.1	LBP の長所	17
3.5.2	LBP 特徴の算出.....	17
3.6	おわりに.....	17
第 4 章	統計的学習手法	18
4.1	はじめに.....	18
4.2	統計的学習手法による物体検出.....	18
4.3	SVM.....	20
4.3.1	マージンの最大化.....	20
4.3.2	非線形 SVM.....	21
4.3.3	HOG 及び SVM による顔検出実験	22
4.4	AdaBoost	23
4.4.2	AdaBoost による強識別器の作成	23
4.4.3	弱識別器を選択するルール[30]	23
4.4.3	エラー率の算出	24
4.4.1	AdaBoost の特徴	24
4.4.5	Cascade.....	25
4.5	ニューラルネットワーク	26
4.5.1	基本構造	26
4.5.2	ニューロンのモデル化について.....	27
4.5.2	ニューロンの学習について.....	29
4.5.3	誤差逆伝播法[34].....	29
4.6	おわりに	30
第 5 章	評価実験	31
5.1	はじめに.....	31
5.2	実験内容	31
5.2.1	実験目的	31

5.2.2	実験環境	31
5.2.3	実験データ	32
5.2.4	評価標準	35
5.3	実験手法	35
5.4	実験結果	35
5.5	考察.....	38
5.6	おわりに.....	39
第 6 章	総括	40
6.1	まとめ.....	40
6.2	今後の課題	41
6.2.1	マンガ検出用学習サンプルデータベースの構築.....	41
6.2.2	新世代検出アルゴリズムの検討	41
謝辞	42
参考文献	43
図目次	46
表目次	48
研究業績	49

第 1 章

序論

1.1 研究の背景

近年、マンガは電子書籍市場における重要なコンテンツである。電子書籍の発展に従い、スマートフォンやタブレットなどのモバイル端末でマンガを読む人が多くなっている。スマートフォンなどの端末では、搭載される画面サイズおよび画面解像度の制限から、コミック上のセリフなどの文字の可読性が低くなっていると考えられる。デジタルコミックは、従来の紙媒体のコミックと異なり物理的な制約がないため、従来のコミックの枠にとらわれな表現や利用が可能になる。そのためには、何らかのメタデータを画像から抽出する必要がある。しかし多くの場合 JPEG などの画像ファイルとしてページ単位で与えられているため、その画像の中からコミックを構成する要素を取り出す技術が必要になる。モバイル端末に対して多くのアプリケーションが開発され、ユーザーは単純に読むだけでは満足しなくなり、マルチメディア付加機能の持つサービスを期待している。例えば、ユーザーに対応するキャラクター登場シーンの検索機能や人気キャラクターセリフ検索機能などである。コンテンツプロバイダーは以上のような便利性の高いサービスの提供を目標としている。こうした要求に応える技術として、画像処理分野を中心に、コミックの画像ファイルを対象としたコマの識別やキャラクター、吹き出しの抽出などに関する研究が様々な進められている。現在、紙面マンガの総数は非常に多く、市場の発展に従い、電子コミックのコンテンツ数も

増加していくと想定できる。しかしながら、キャラクター検出や認識及び検索は人手で行われており、これらの処理の自動化に関して社会的な要求があるといえる。

一方、インターネットの普及に伴い、Blog やコンテンツ共有サービスを利用して自らが描いたイラストやマンガを公開し、他者に閲覧、評価してもらうことができるようになってきている。こうした状況により、初心者であってもコミックを制作できるように支援する技術に注目が集まっている。既に、“コミ Po!”[1]のような、事前に用意されたキャラクターの顔や体、オブジェクト等を組み合わせることにより、絵を描くことなくコミックを生成できる商用のマンガ作成支援システムが登場した。コミック作成者支援技術に共通するのは、過去に上梓されたコミックから取得した知識や事前に用意されたプリミティブ(キャラクターやオブジェクト、背景など)を利用している点にある。したがって、特にマンガにおけるキャラクター認識が重要であると考えられる。

コミック画像は主に線で描かれる。その上にトーンや塗りつぶし等の手法により陰影づけされたキャラクターや風景などの絵の部分と、吹き出し、セリフや擬音語などの文字要素が複雑に配置されている。

コミックに登場するキャラクターの識別手法として、HOG 特徴量(Histograms of Oriented Gradients)を手がかりにして画像内の顔候補を特定し[2]、その顔候補と予め作成したキャラクターの顔画像データベースとのマッチングを行い、顔候補画像がどのキャラクターであるかを識別する手法が提案されている。

1.2 研究の目的

コミックのキャラクター検出に関する研究として、石井らは輝度に注目する HAAR-LIKE 特徴量及び SVM を利用しキャラクターの顔を検出する手法を提案している[3]。また、石井らはエッジ情報が抽出できる HOG 特徴量及び SVM を用いてキャラクターの顔と目を検出する手法を提案している[4]。しかしながら、これらの手法では、キャラクターの検出成功率があまり高くなく、約 60%である。石井らの手法では、コミック画像からキャラクターの正面顔を比較的正確に検出できるが、HOG 特徴を用いて捉えるエッジ情報の性質から、顔の向きに対する安定性が弱いという問題点がある。この場合には、すべてのキャラクターに対して検出結果が不十分である。コミック画像のキャラクター検出を行う際に、キャラクター画像を学習する必要がある。石井らのキャラクター検出手法では、キャラクターの顔であ

るポジティブ画像を事前に若干枚用意し、キャラクターの顔ではない画像ネガティブ画像をポジティブ画像枚数の 2 倍使用している。学習画像はすべて手で切り出したものである。石井らは学習画像を一定サイズに正規化させ、機械学習アルゴリズム SVM を用いるポジティブ画像とネガティブ画像を別々に学習させている。次に、学習結果による特徴量を計算し、未知画像に対して検出実験を行っている。この手法では、マンガの作者により、マンガのスタイルが完全に変わるため、検出能力に限界がある。さらに一部マンガの場合、キャラクターの登場シーンは多くないため、学習画像の枚数が制限される。一方、一般画像の顔検出を行う場合には、数万枚の学習画像を準備する必要がある。したがって、マンガに対して少ない学習画像を用いてそのまま検出実験を行ってしまうと最終的に期待される検出成功率が得られないという問題がある。

本研究では、より高い検出成功率を得るために、マンガ画像に対する従来の顔検出特徴量の性能を評価し、それぞれの特性について検討する。

1.3 本論文の構成

以下に本章以降の本論文の構成を示す。

第 2 章 本章では、近年の電子コミックマーケットの動向や、その閲覧端末について述べる。

第 3 章 最近までに提案されてきている顔検出に適した特徴量について述べる。例として、HAAR-LIKE 特徴量, SIFT 特徴量, HOG 特徴量及び LBP 特徴量について述べ、コミック画像に対する顔検出効果について検討する。

第 4 章 ターゲット検出を目的とした統計的学習手法について述べる。SVM, AdaBoost 及びニューラルネットワークについて詳しく説明しながら、各アルゴリズムはマンガ画像との相性を検討する。

第 5 章 第 3 章で述べた特徴量と第 4 章で述べた統計的学習手法を用いて、マンガ画像におけるキャラクターの顔検出比較実験を行った結果を示して検討する。

第 6 章 本論文を総括して、今後の課題を検討する

第 2 章

電子コミックの現状

2.1 はじめに

本章では、近年の電子コミックマーケットの動向や、その閲覧端末について述べる。はじめに電子コミックマーケットの市場規模の変動やコンテンツ配信ビジネスの形態について述べ、次に電子コミックの将来展望について述べる。

2.2 電子コミックの発展史

世界初のデジタルコミックは寺沢武一が 1985 年に発表した『黒騎士バット』である。「デジタルコミック」という用語自体も寺沢による造語で、1987 年頃から使い始めたとしている [5]。1990 年代半ばより、既存のマンガ作品の画像を取り込んだり、通常のマンガと同じようなコマ割り構成で新規に描かれた作品に、BGM や効果音、声優による音声などを加えた作品が登場した。ボタンを押すことで読み進めていくため、マンガを再現するという点において本質的な意味でのデジタルコミックと言える。1990 年代後半からは、インターネットの普及によりマンガのオンライン配信が可能となり、これらもデジタルコミックと呼ばれている。作品は紙媒体上で発表されたもののデジタル化、およびオリジナル作品に大別され、趣味や同人活動の一環で配信するものはウェブコミックと呼ばれることもある。配布フォーマットは Adobe Flash, XMDF, PDF などで、紙媒体で発表するより費用を抑えられるというメリットがある。近年は携帯電話の普及および高性能化により、携帯電話で閲覧できる携帯コミックも発表されている [6]。2004 年 8 月には NTT 西日本系のコンテンツ配信会社 NTT ソルマーレが「コミック i」のサービス [1] を、翌 05 年 5 月には「コミックシーモア」

のサービス[1]を開始して、携帯電話で読む電子コミック、つまり携帯コミックが定着する。2005年に46億円だった携帯コミックの市場は2009年には513億円を突破する。携帯コミックは第3世代(3G)機の普及とパケット定額制によって読者を増やした。また、電話の利用料に乗せてコンテンツの利用料を徴収するキャリア課金は、配信元にとっての利便性を上げた。カード決済やプリペイドカードによる決済よりも、簡便で確実な売上回収が可能となった。図(2.1)は現在スマートフォンで人気のマンガの閲覧アプリケーションを示す。



図 2.1 現在スマートフォンで人気のマンガの閲覧アプリケーション図

2.3 電子書籍マーケットの現状

2013年から2014年の一年間に電子書籍ストア間の熾烈な競争が始まった。より多様なビジネスモデルが探られ、各ストアの広告宣伝活動も積極的に行われた。その結果、電子書籍そのものの認知が広がりユーザー数が着実に増加し、電子書籍と電子雑誌を合わせた電子出版市場は1,000億円を突破し本格的な拡大期となった[8]。

2.3.1 電子コミック市場

2008年6月にはアップルの社の多機能のスマートフォン、iPhone 3Gが日本国内で発売された。しかし、キャリアがソフトバンク1社だったこともあって、それまでの携帯電話の市場は当面安泰だと考えられていた。携帯配信各社もスマートフォン向けを意識しながらも、携帯コミックが中心という姿勢を崩さなかった。積極的にiPhone向けのマンガ配信を進めたのは、イーブックジャパンなど、PC向けにマンガを配信してきた電子書店各社や既存の出版社であった。2012年度から再び電子コミック市場は拡大を始めた。携帯コミックは縮小したが、スマートフォンやタブレット端末向けの配信がそれ以上に拡大したためである。「電子書籍ビジネス調査報告書」(インプレス R&D, 2014)によると、電子書籍市場規模は2013年度で前年比33.5%増の936億円である。2年前の予想と比べ、スマー

トフォンの市場の伸びが大きく、フィーチャーフォンの市場の減少が顕著となった。2013年度は、マンガ市場がスマートフォン向け電子書籍の中心となっている。2012年の資料[8]ではスマホ及びフィーチャーフォン向け電子書籍コンテンツの内訳を見ると、マンガを中心としたモバイル向け電子書籍市場が、2012年度は574億円と、電子書籍市場の78.7%を占めていることが分かる。インプレス総研の報告書によれば、5年後の2018年には電子書籍の国内市場は3000億円に達すると予測されている。このままの比率が維持されるならば、電子コミック市場の規模は2400億円となり、現在のマンガ雑誌と単行本を合わせた市場規模である3669億円のおよそ3分の2にまで拡大するとみられている。マンガ雑誌、単行本(コミックス)の市場は毎年110億円程度の減少を続けているため、10年以内には紙と電子の逆転劇が起きると予想される。図(2.2)は電子書籍市場の割合と規模推移を示す。

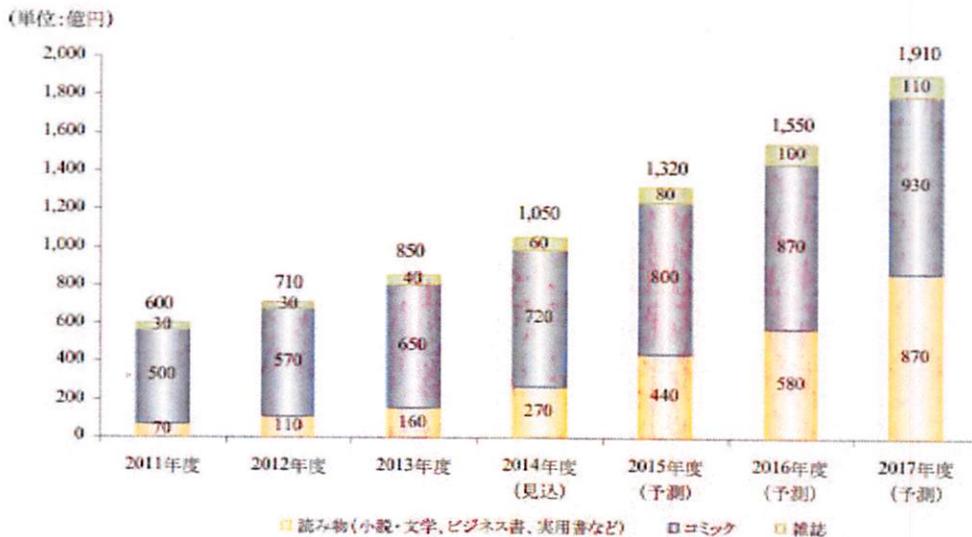


図 2.2 電子書籍市場の割合と市場規模推移(文献[8]より引用)

2.3.2 紙からデジタルへの転換

2014年8月、NHN PlayArtが運営するスマートフォン向け無料コミック配信サービス「comico」の人気作品『ReLIFE』[1]の単行本がアーススター エンタテインメントから発売された。電子コミックは、電子書籍端末やタブレット端末で読むことが一般に想定されている。しかし2013年秋にスタートした「comico」はスマートフォンに特化したサービスであり、2

値画像ではなくフルカラーである。また、小さな画面でも読みやすいように縦スクロールで読むスタイルになっているのが特長である。

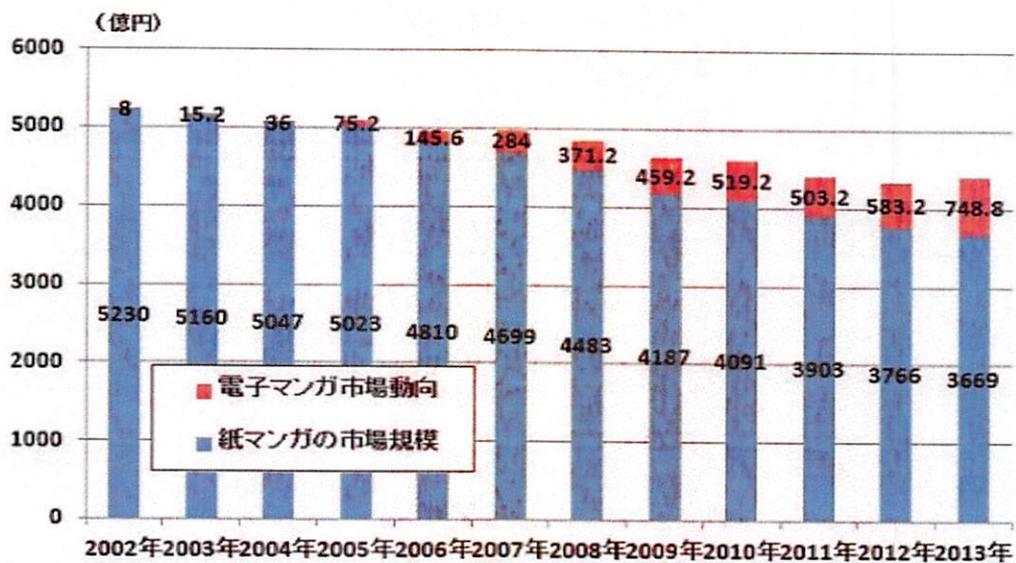


図 2. 3 電子コミックと紙コミックの市場規模対比(文献[8]より引用)

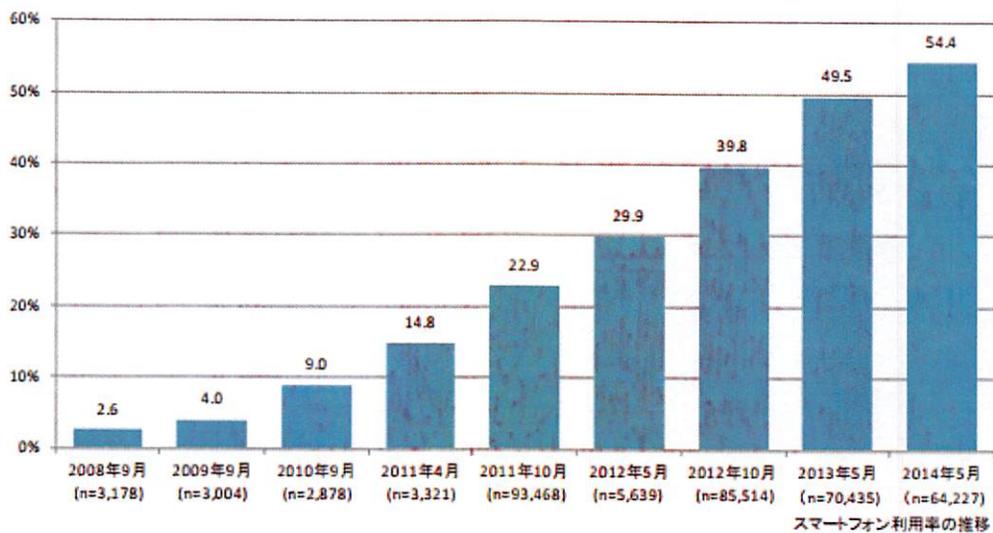


図 2. 4 スマートフォン利用率の推移(文献[8]より引用)

これまでのマンガ作品は、人気雑誌に連載されたものが単行本化されるというパターンが多かった。しかし、「comico」はアプリのダウンロード数 300 万(2014 年 12 月現在)でありながら、完全な電子雑誌である[1]。この傾向が続くと、読者は紙の単行本を買うケースがより低下すると考えられる。

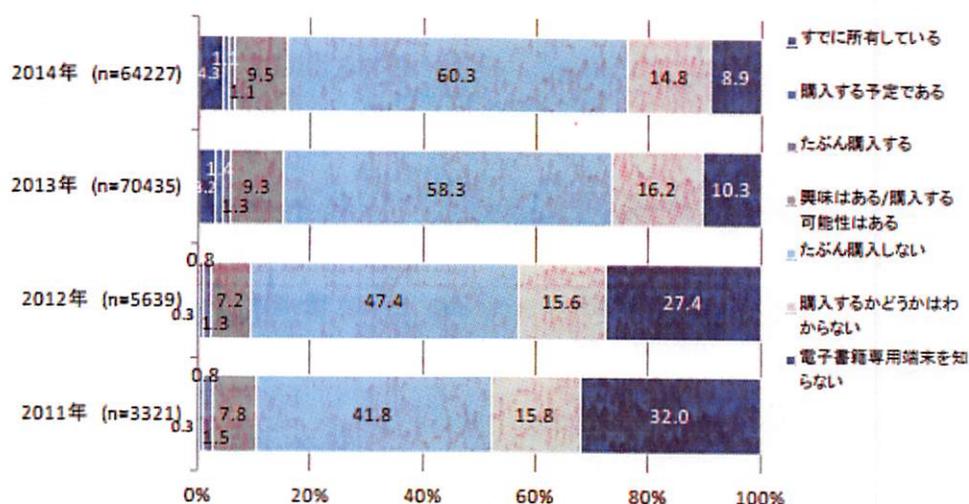


図 2.5 電子書籍専用端末所有率の推移(文献[8]より引用)

今後は、大手出版社も電子コミックを大規模展開していくと考えられる。

2.4 おわりに

本章では、電子書籍市場の近年の変化と、電子書籍の閲覧端末及びアプリケーションに関して述べた。電子書籍市場では、スマートフォンによるサービスの急成長と、電子書籍市場全体に占める割合から、その重要性を示した

第 3 章

画像局所特徴量

3.1 はじめに

本章では、最近までに提案されてきている顔検出に適した特徴量について述べる。顔検出とは、画像中から顔の位置を検出する技術である。具体的に言えば、画像の中から「位置」、「大きさ」、「顔の傾き」、「画像の回転」、「顔向き」、「照明変化」、「隠れ」にかかわらず人の顔の位置と大きさを特定することである。例として、HAAR-LIKE 特徴量、SIFT 特徴量、HOG 特徴量及び LBP 特徴量について述べ、コミック画像に対する顔検出効果について検討した。

3.2 「パターン認識」による顔検出の基本的な流れ

顔検出の目的は、画像中を探索して「人間の顔の映っている領域」を探すことにある。その基本的な処理の流れとしては、対象の静止画像一枚に対し、画像全体に顔の大きさの探索窓(顔検出処理を行う判定領域)を動かし、探索窓が通過する座標ごとに「探索窓内の画像が、顔画像であるかどうか」の判定を行う。こうした探索窓領域ごとの判定処理を画像中すべてに行うことで、最終的に顔の領域だと判定された探索窓の場所を、検出した顔の位置として決定するのが顔検出の処理の流れである。

3.3 HAAR-LIKE 特徴量

顔検出で現在主に使用されているのは、Viola と Jones が考案した「HAAR-LIKE 特徴」という特徴量[9]である。HAAR-LIKE 特徴量を用いた顔検出アルゴリズムでは、顔検出を行いたい対象の 1 枚の画像に対して、探索窓(例えば 8ピクセル×クピクセルのような判定領域)を左上から走査して順番に動かす。この探索窓の領域ごとに、あらかじめ Boosting という機械学習アルゴリズムにより作成しておいた Cascade 構造の識別器により、顔画像であるかどうかの判定を行う。このように、「各探索窓において顔が存在するかどうかを Cascade 構造の識別器により高速に判定する」というのが、顔検出の基本的な処理フローである。

3.3.1 局所的な明暗差の集合

図(3.1)のように、人間は生まれた直後でも、目と口を表す 3 つの黒い点に注目することが知られている。全体を見なくても、特徴点な部分を見れば顔とわかる。細かな特徴を見なくても、大まかな明暗差で顔とわかることが知られている。

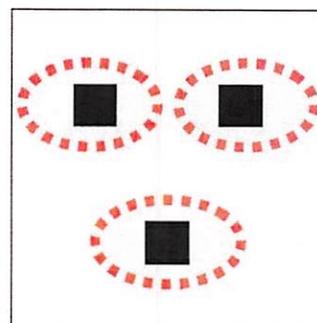


図 3. 1 人間の顔のパターン(文献[10]より引用)

人間の顔は、実は局所的な明暗差の集合である。図(3.2)のように、大量のサンプルから局所領域の集合を学習により、複数の明暗差を捉えることで顔であるか否かを判定可能と考えられる。

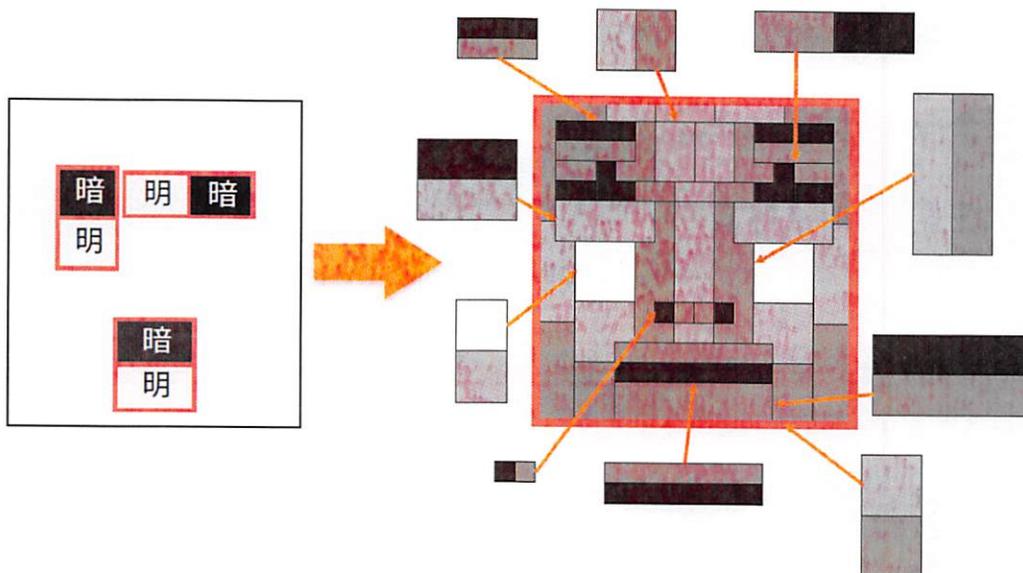


図 3.2 顔の明暗差分析図(文献[10]より引用)

3.2.2 HAAR-LIKE 特徴パターンの分類

一般的に使用される HAAR-LIKE 特徴パターンは、図 3.3 に示すように、線形特徴、エッジ特徴及び中心特徴に分けられる。

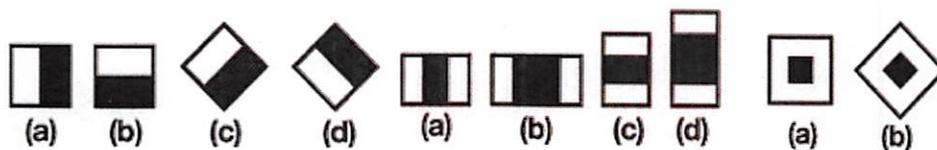


図 3.3 エッジ特徴(左) 線形特徴(中) 中心特徴(右)

3.2.3 HAAR-LIKE 特徴の計算[11]

各特徴値を計算するには黒い領域の各画素の輝度値の和と白い領域の各画素の輝度値の和の差を求める必要がある。計算されるこの差は HAAR-LIKE 特徴の値となる。この差は式(3.1)で算出できる。

特徴量 $H(r1,r2)$ = AとBの平均輝度の差を算出

$$\begin{array}{c}
 \begin{array}{|c|c|}
 \hline
 \text{A} & \text{B} \\
 \hline
 \end{array} \\
 \hline
 \end{array}
 \quad
 H(r1,r2) = \frac{S(r1)}{\text{領域Aの平均輝度}} - \frac{S(r2)}{\text{領域Bの平均輝度}} \quad (3.1)$$

3.2.4 HAAR-LIKE 特徴の積分画像

HAAR-LIKE 特徴のパターンの各部分輝度値を計算すると、矩形領域の面積と同じ回数
の加算が必要であるので、積分画像による高速化することができると考えられる。

学習サンプルは数多く必要であり、さらに特徴矩形の数は非常に多く、特徴量を計算す
るために多量の矩形内の画素の和を計算しなければならない、そのため学習と検出のス
ピードが大幅に遅くなる。そこで、積分画像による高速化が提案されている。矩形の特
徴量を計算する場合、この特徴矩形の端点の積分画像だけにより求めることができる。した
がって、特徴矩形のサイズが変わっても、非常に少ない計算回数で特徴量を求めること
ができる。図 3.4 のように、事前に積分画像の値が求められていれば、任意矩形の全
ての画素の輝度積分が簡単に計算できる。矩形領域 D の全ての画素輝度積分は矩形端
点の積分画像値で得られる。

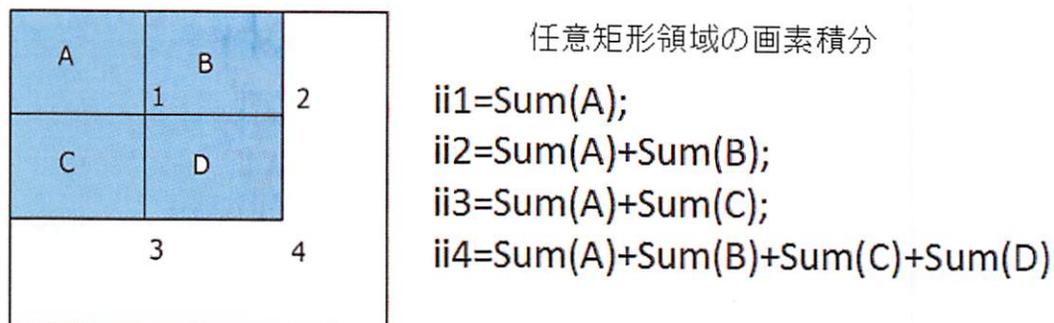


図 3.4 任意矩形領域の画素積分の計算式

3.3 SIFT 特徴量

Scale-Invariant Feature Transform (SIFT) は、特徴点の検出と特徴量の記述を
行うアルゴリズムである[12]。画像の回転・スケール変化等に頑健な特徴量を記述するた
め、画像のマッチングや物体認識・検出に用いられている。SIFT 特徴は 1999 年に
D.G.Lowe によって提案され、2004 年前後に成熟した技術となっている。

3.3.1 SIFT の長所

- 1) SIFT 特徴は画像の局所特徴であり, 画像の回転, スケール変化, 照明変化等に頑健な特徴量を記述することができる[13].
- 2) 独自性(Distinctiveness), すなわち豊富な情報量を持ち, 大規模なデータベースで, 高速かつ正確なマッチングに適性が高い.
- 3) 多量性, すなわちいくつかのオブジェクトは, SIFT 特徴ベクトルを大量に生成することができる.
- 4) 高速性, すなわち最適化された SIFT マッチングアルゴリズムはリアルタイムでマッチングを達成することができる[14].
- 5) 拡張性, すなわち他の特徴と簡単に組み合わせることができる.

3.3.2 SIFT 特徴で処理できる問題[15]

- 1) 目標回転, 拡大縮小, 平行移動.
- 2) 画像のファイン/投影変換 (viewpoint)
- 3) 照明影響 (illumination)
- 4) オブジェクト遮る(occlusion)
- 5) 乱雑なシーン(clutter)
- 6) ノイズ

3.3.3 SIFT 特徴とコミック画像の親和性

マンガ画像では, 線画, トーン部分の双方において, SIFT のキーポイントになりうるエッジ成分が無数に存在する. このため, 画像中のあらゆる箇所に SIFT のキーポイントが検出されてしまう. これは, 完全に同一の画像や部位を照合する目的には使用できるが, 変化のある特定の登場人物などを見つけ出す用途にはキーポイントの制御が困難である. また, 多数のキーポイントから有意な情報のみを取り出すことは難しい. マンガ画像上で SIFT 特徴量取得した結果例を図(3.5)に示す. 無数のキーポイントが検出されている事が確認できる[16].

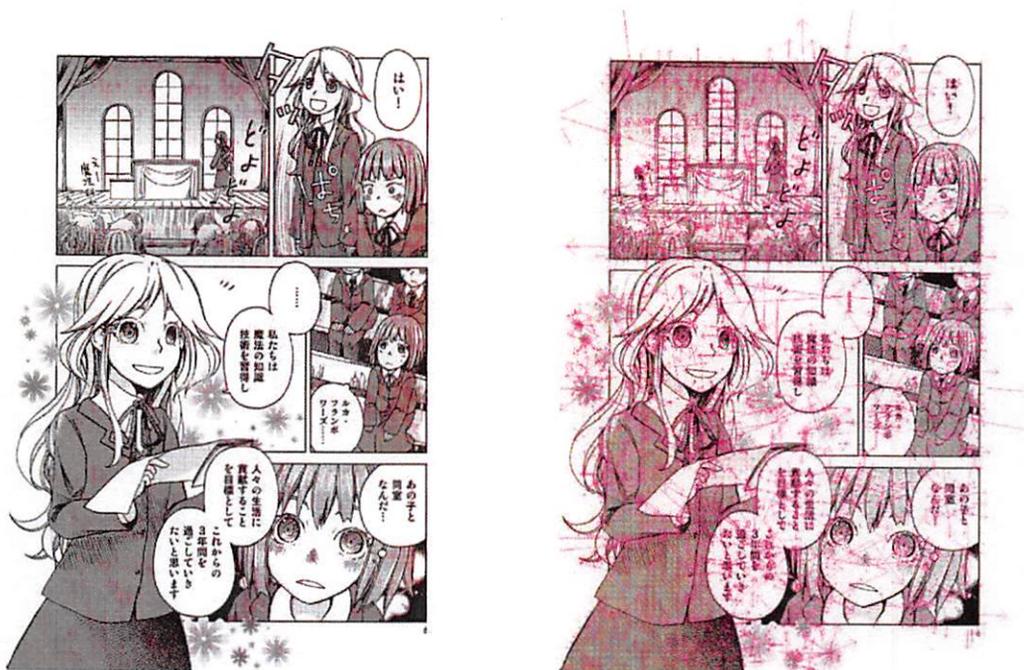


図 3.5 マンガ画像例と漫画画像上の SIFT 特徴量(文献[16]より引用)

3.4 HOG 特徴量

Histograms of Oriented Gradients (HOG)特徴量は, 入力画像の勾配(微分画像)を求め, それを局所領域ごとに区分して勾配方向のヒストグラムを取ったものである[17]. 輝度差が捉えられなくても顔の形状を解析の手掛かりとして利用できる. また, マンガ画像の場合, 顔の向きと形状の種類が多いため, HOG は幾何学的変換に強く対応できると考えられる[18].

3.4.1 HOG 特徴の算出アルゴリズム

1) 輝度の勾配方向と勾配強度の算出[19]

エッジ情報はエッジ方向, エッジ強度を式(3.2)-(3.4)で算出する.

$$\text{勾配強度} : m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (3.2)$$

$$\text{勾配方向} : \theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (3.3)$$

$$\begin{cases} f_x(x, y) = L(x+1, y) - L(x-1, y) \\ f_y(x, y) = L(x, y+1) - L(x, y-1) \end{cases} \quad (3.4)$$

2) セル領域分割

先のステップで作成したエッジ画像はピクセル単位の特徴のため形状変化に非常に弱い。そこで、画像を 5 ピクセル四方の正方形領域に分割する。この分割された領域をセル領域と呼ぶ。一つのセル領域には 25 ピクセル分のエッジベクトルが含まれている。

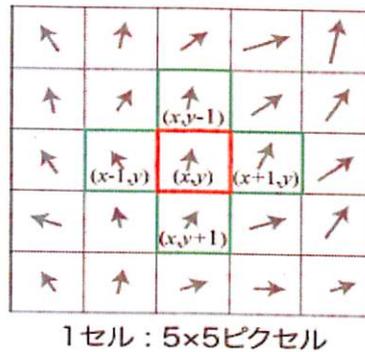


図 3. 6 1セルの例(文献[19]より引用)

3) ヒストグラムの作成

先のステップで分割した各セル領域に含まれるピクセルのエッジ方向、エッジ強度から、ヒストグラムを作成する。0° のから 180° までを、20° をずつ 9 方向に分割する。

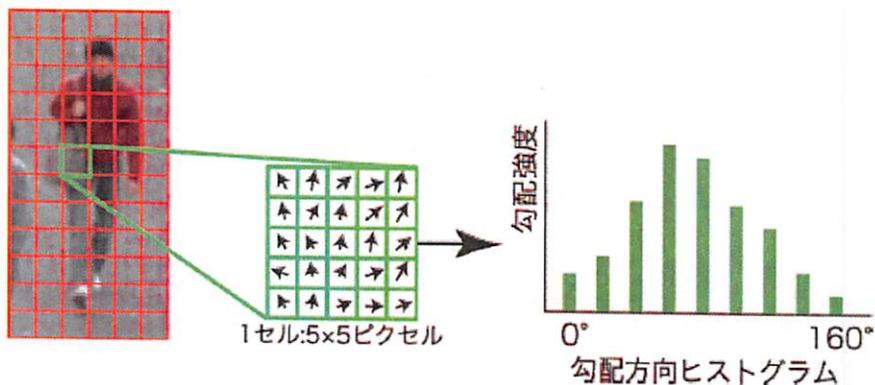


図 3. 7 HOG 特徴のヒストグラムの例(文献[19]より引用)

エッジ情報を局所領域でヒストグラム化することによって、物体の形状変化に頑健な特徴量を得ることが出来る。これは、セル内でエッジが平行移動した場合でも、ヒストグラム

の値は変化しないことに起因しており、セルの大きさの範囲で、物体の形状が変化しても、HOG 特徴量はほぼ同一のヒストグラム値を持つことが出来る。

4) ブロック領域による正規化

3セル×3セルの領域をブロック領域と呼ぶ。このブロック領域には9個のヒストグラムが存在し、この9個のヒストグラムを用いて各セルのヒストグラムを正規化する。正規化を行ったブロック領域を1セルずつスライドさせ、各セルは複数回正規化が行われる。

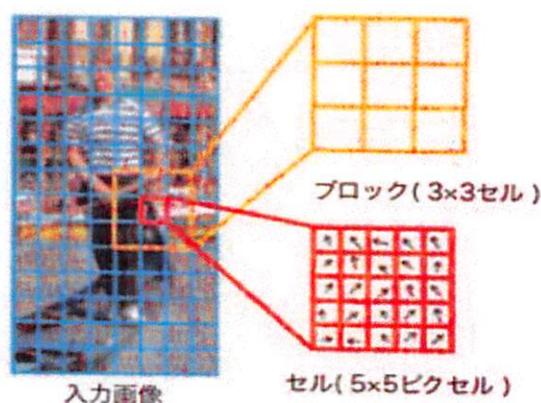


図 3.8 1 ブロックの例(文献[19]より引用)

3.4.2 HOG 特徴の次元数算出

HOG 特徴量は非常に大きな特徴量であり、最終的な HOG の次元数 V は、ヒストグラムの次元数を v 、画像の横幅を w 、縦幅を h 、セルサイズを c 、ブロックサイズを b としたとき、式(3.5)で表される。

$$V = (w/c - b + 1)(h/c - b + 1)b^2v \quad (3.5)$$

例: 250x250の入力画像の場合

$$\bullet V = (250/5 - 3 + 1) * (250/5 - 3 + 1) * 3 * 3 * 9 = 186624$$

3.5 LBP 特徴

LBP (Local Binary Pattern)では、T. Ojala (1994)らによって提案された特徴量である[20]。通常 3×3 の画素領域で計算され、画像の局所的な特徴を抽出できる。特

に、画像の照明変化の影響を受けにくい。そして、高速に計算できるという特長がある。注目画素と近傍画素の輝度の大小関係を表現することによって、局所領域におけるバイナリパターンをヒストグラム化する。局所的な輝度の分布をバイナリコードとして捉える。

3.5.1 LBP の長所

LBP 特長量は濃度変化に対して不変な特徴量である。さらに、計算コストが少ない。

3.5.2 LBP 特徴の算出[21]

LBP の計算には 3×3 画素領域内の画素値が使われ、入力画像の階調数に関わらず LBP 画像 は 256 階調の画像になる。まずは、図(3.9)のように、256 階調グレースケール画像へ正規化される。次に、画像を複数のブロックに分割し、ブロック内の画素に対して LBP を計算する。そして、ブロック毎に LBP のヒストグラムを生成する。最後に、ブロック数分のヒストグラムを連結して特徴を表す[22]。

例えば、注目点の画素値 $I(r)$ と、その 8 近傍にある画素値 $I(n)$ を比較すると、

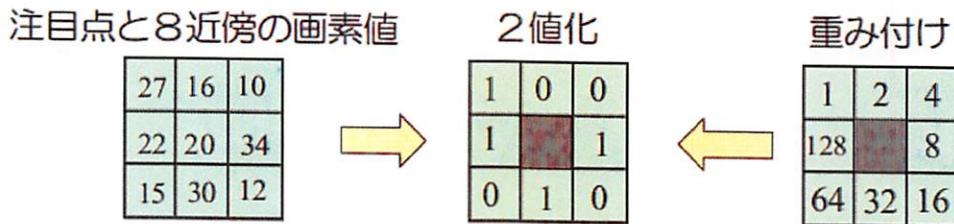


図 3.9 LBP 特徴の算出手法(文献[21]より引用)

$$\begin{cases} 1 \dots \text{if } I(n) \geq I(r) \\ 0 \dots \text{if } I(n) < I(r) \end{cases}$$

LBP Pattern=10101001

LBP value=128+32+8+1=169

3.6 おわりに

本章では、HAAR-LIKE, SIFT, HOG 及び LBP 特徴を例として、近年で提案された顔検出向きの特徴量について述べる。そして、各特徴の算出方法についてそれぞれ紹介した。次の第 4 章では、特徴量と組み合わせて統計的学習手法について述べる。

第 4 章

統計的学習手法

4.1 はじめに

本章では、ターゲット検出を目的とした統計的学習手法について述べる。SVM, AdaBoost 及びニューラルネットワークについて詳しく説明しながら、各アルゴリズムはマンガ画像との相性を検討した。

4.2 統計的学習手法による物体検出

統計的学習手法による物体検出の処理フローを図(4.1)に示す。まず、大量な学習サンプルを用意する必要がある。人物検出を目的とした場合、人の画像を学習ポジティブ画像として設定し、人以外の学習サンプルを学習ネガティブ画像とする必要がある。次に、用いるべき複数の特徴量を抽出し、識別器に入力する。本章の実験では HOG 特徴量を用いる。最後に検出ウィンドウをラスタスキャンした後、識別結果を統合する。

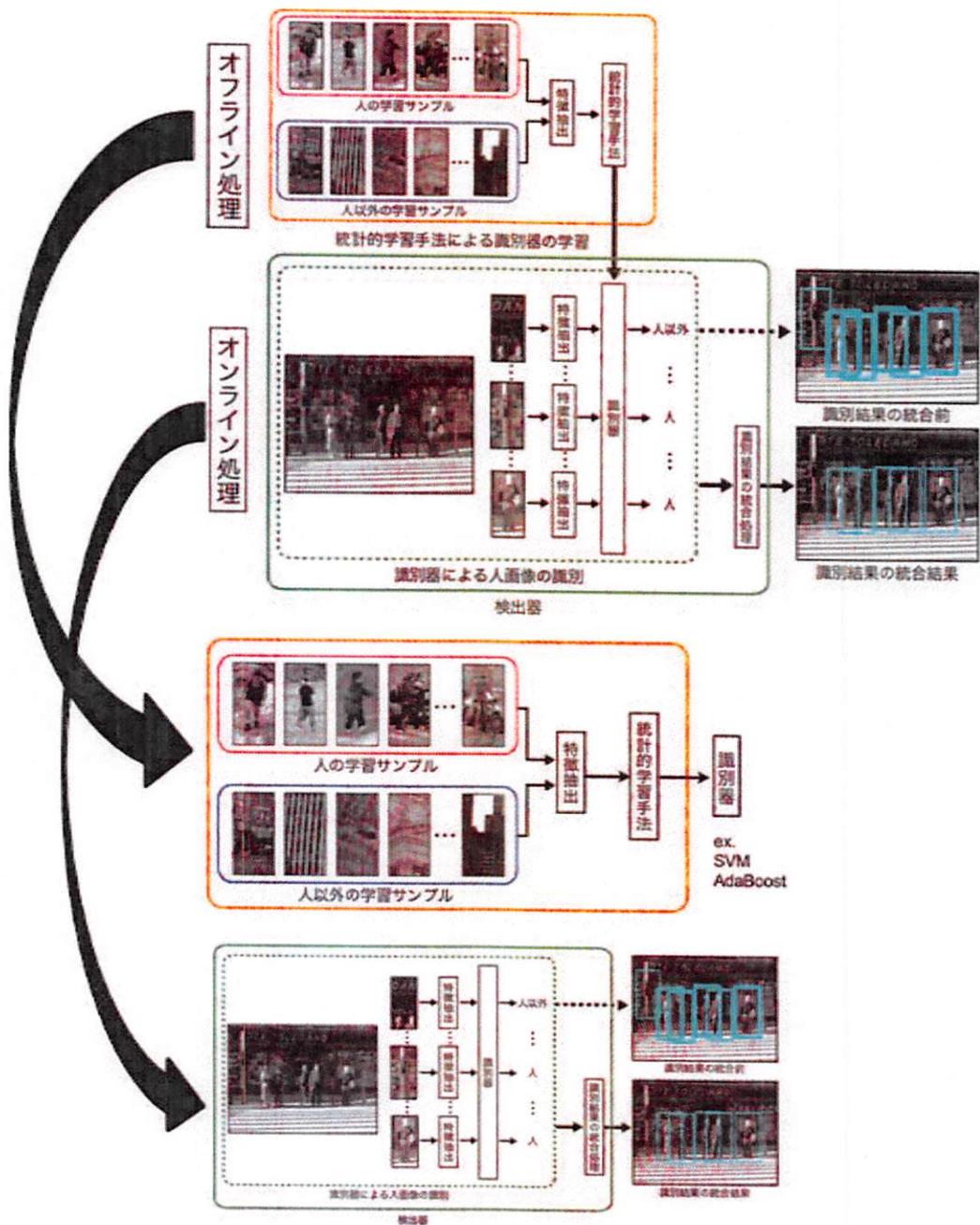


図 4.1 一般的な統計的学習手法による物体検出の流れ(文献[15]より引用)

4.3 SVM

サポートベクターマシン(Support Vector Machine, SVM) は, 教師あり学習を用いるパターン認識モデルの一つであり, 識別や回帰分析で用いられている[23].

サポートベクターマシンは, 現在知られている手法の中でも認識性能が優れた学習モデルの一つである. サポートベクターマシンが優れた認識性能を発揮することができる理由は, 未学習データに対して高い識別性能を得るための工夫が含まれることによる.

4.3.1 マージンの最大化

顔と顔でない画像に分離する場合, 「識別線の引き方」に注意する必要がある[24]. まず, 2次元の特徴空間に図(4.2)のような二つのクラス A と B に属する学習サンプルがあるとす. これをうまく分離するために, SVM では「マージン最大化」という処理を行う.

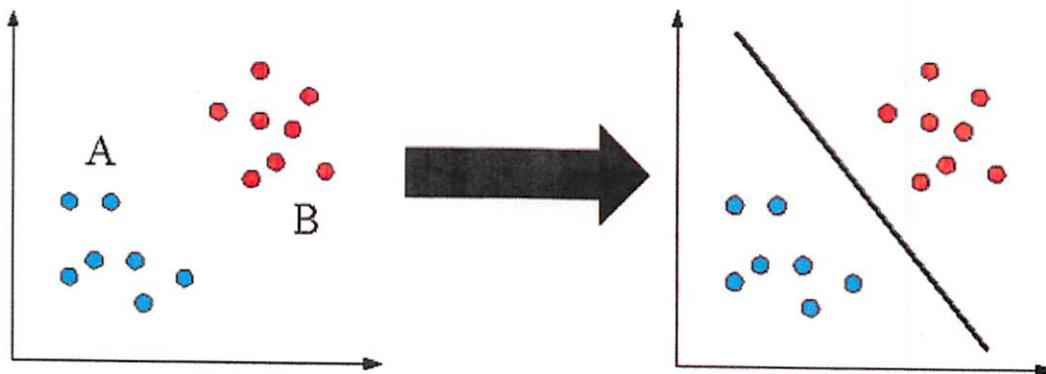


図 4. 2 識別の例(文献[24]より引用)

最適な識別境界を得る手法が大きく分けて二つ存在する. 一つはパラメトリックな方法, つまり統計的に学習データの分布を考えて識別境界を引く方法である. もう一つはノンパラメトリックな方法, つまり与えられた学習データをすべて正しく識別できるようにする方法である.

SVM は後者のノンパラメトリックな手法である. バックプロパゲーション学習もノンパラメトリック手法であるが, SVM がそれと異なるのは識別境界の位置を決定する明確な基準を持っていることである.

SVM では、学習データの中で最も他クラスと近い位置にあるもの(これをサポートベクトルと呼ぶ)を基準とし、そのユークリッド距離が最も大きくなるような位置に識別境界を設定する[25]。クラスの最端から他クラスまでのマージンを最大にするためマージン最大化と呼ばれる。

図(4.3)のように、マージン最大化の基準にはユークリッド距離が用いられている。このような明確な基準を与えているということは、ノンパラメトリックな手法では他に例がなく、SVM の最も優れた部分である。

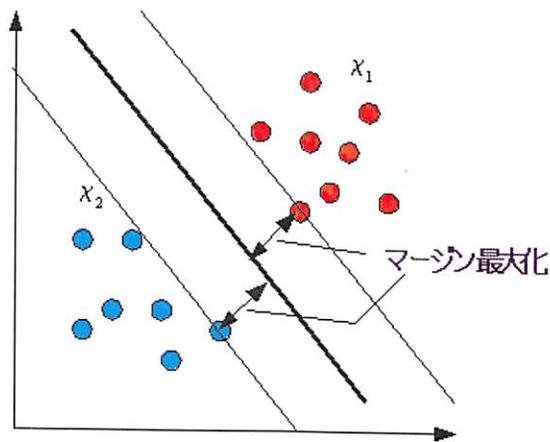


図 4. 3 SVM によるマージン最大化処理(文献[24]より引用)

4.3.2 非線形 SVM

1963 年にウラジミールヴァプニクが発表した初期のサポートベクターマシンは、線形分類器にしか適用できなかった。しかし、図(4.4)のように、再生核ヒルベルト空間の理論を取り入れたカーネル関数を用いてパターンを有限もしくは無限次元の特徴空間へ写像し、特徴空間上で線形分離を行う手法が 1992 年に ヴァプニクらによって提案された。これにより、非線形分類問題にも優れた性能を発揮することがわかり、近年特に注目を集めている[26]。

なお、カーネル関数を取り入れた一連の手法では、どのような写像が行われるか知らずに計算できることから、カーネルトリック (Kernel Trick) と呼ばれている。顔検出の場合には線形 SVM を用いる。

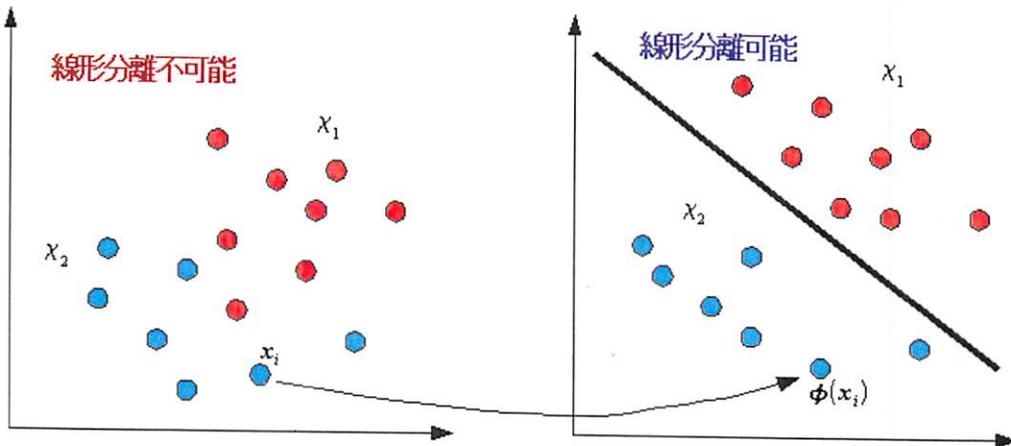


図 4.4 線形と非線形 SVM の例(文献[24]より引用)

4.3.3 HOG 及び SVM による顔検出実験

ここでは, HOG 特徴量及び SVM 学習アルゴリズムを用い, マンガ画像における顔検出実験を行った. 実験結果は表(4.1)に示す.

表 4.1 HOG 及び SVM によるによる顔検出実験

キャラクター	既知画像		未知画像	
	検出率(%)	誤検出率(%)	検出率(%)	誤検出率(%)
A	97.5	12.6	71.0	44.5
B	96.2	11.5	68.5	46.2
C	93.3	9.9	66.3	51.1

ここでは,

$$\text{検出率} = \frac{\text{成功検出の数}}{\text{検出された画像すべての顔の数}} \quad (4.1)$$

$$\text{誤検出率} = \frac{\text{誤検出の数}}{\text{すべての検出の数}} \quad (4.2)$$

表(4.1)によると, 既知画像に対する検出結果が未知画像よりかなり良い結果を得た. 特に, 未知画像の場合, 誤検出率が大幅に上がってしまった. マンガは白黒二色で構成される 2 値画像である. 主な要素は点, 線で描かれるキャラクター, 背景とテキストであるので, エッジ情報を基にした HOG 特徴量によって検出を行うと, マンガ画像の中で検出に影響を与える要素が多数存在するため誤検出率が高くなっている.

4.4 AdaBoost

AdaBoost は, 1997 年に Y.Freund 等が提案した統計的学習手法 Boosting(強化学習)の一つである[27]. 探索窓領域において顔かどうかを判定する Cascade 構造の識別器は, Boosting という機械学習アルゴリズムにより事前に作成する. Boosting では, 認識率はある程度低いがシンプルで計算量が少なく済む識別器である「弱識別器」を用いる. 各弱識別器が誤って識別したものをそれぞれ担当して正しく識別するように学習させていき, それら弱識別器を連結させて用いることで, 弱識別器同士の弱点を補いあつた一つの強力な識別器である「強識別器」を構成する[28].

Adaboost は学習サンプルの重みをリアルタイムで適応的に更新できる機械学習アルゴリズムである. 計算量の増加と汎化性能の低下を抑制しながら, 誤識別したサンプルに対し重みを増加する特徴を持っている. AdaBoost アルゴリズムの最終識別器は, 弱識別器の重み付き多数決による処理を行っている.

4.4.2 AdaBoost による強識別器の作成

顔検出では Adaboost アルゴリズムにより, 強識別器を作成する. 複数用意してある弱識別器を順番に学習していき同時にその学習された各弱識別器の全体に対する重要度を決定する. 各弱識別器の学習が終わるごとに, 誤識別された学習データの重みを増加させる.

誤識別したデータの重みを増加させておくことで, 毎回次の弱識別器を学習する際に, 一つ手前の弱識別器が不得意としていた学習データを識別してくれる弱識別器を学習することになる. こうして, 各弱識別器がお互いの苦手なデータの識別を補完し合い, 全体で一つの高精度の強識別器を作ることができる.

4.4.3 弱識別器を選択するルール[30]

- 1) 入力データに対して, +1, もしくは-1 を出力する 2 値関数である
- 2) 弱識別器の正解率は, サンプルの重みを利用して計算する
- 3) 学習サンプルに対して, 正解率が 50%以上である

弱識別器を選択するルールを図 4.5 に示す.

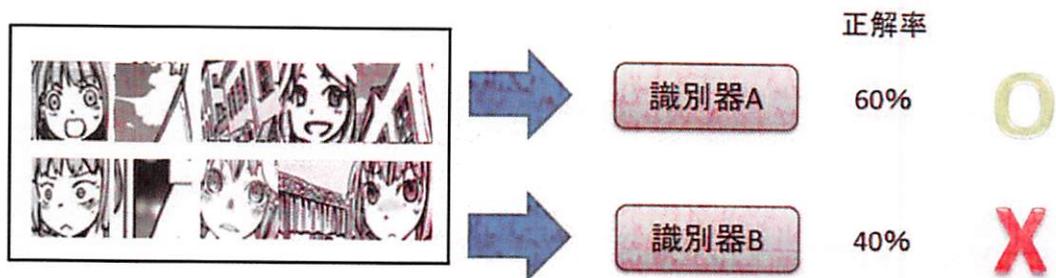


図 4.5 弱識別器を選択するルール(文献[39]より引用)

4.4.3 エラー率の算出

式(4.3)より弱識別器に対するエラー率 $e_{t,m}$ を算出できる. 弱識別器 h の応答値とクラスラベルが異なった場合サンプルの重みを足し合わせる.

$$e_{t,m} = \sum_{i: h_t(x_i) \neq y_i} D_t(i) \quad (4.3)$$

エラー率が 0.5 を超えたら, 弱識別器のパリティ p を反転する.

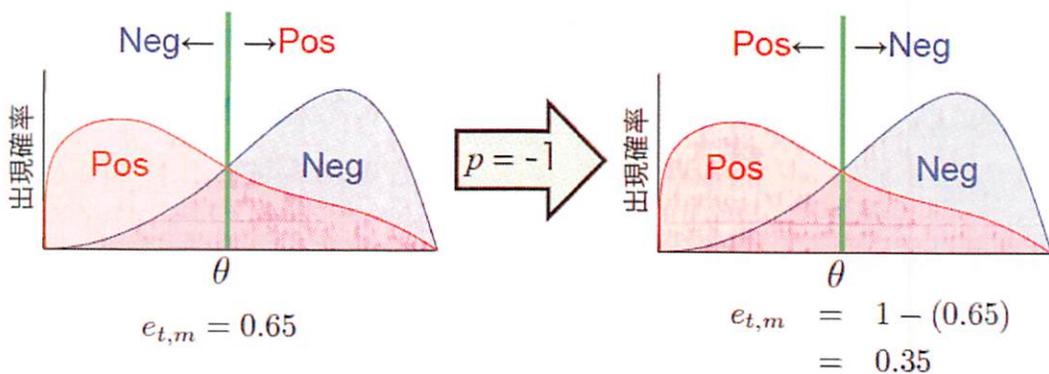


図 4.6 エラー率を算出の例(文献[29]より引用)

4.4.1 AdaBoost の特徴

1) 識別精度が高い

汎化性能に対する理論的な裏付け, 学習後期では識別の困難なデータに特化できる [30].

2) 解析が容易

特徴選択の傾向などを後から調査になるタイプである。

4.4.5 Cascade

実際の検出においては、検出対象の多くは背景などでネガティブに分類される。従って、効果的な少ない特徴量で negative の検出をより早い段階で得ることができれば、検出速度は大幅に向上する。そこで、いくつかの識別器で下図のようなカスケードを作成し、いずれかのステージで negative が検出されれば、false を返す検出器を構築する。

各ステージは AdaBoost の分類器学習で作成し、デフォルトでは学習データのエラー率が最小となるように設定されている閾値(重み)を、検出もれが最小になるように設定する。カスケードの学習において、

- 1)ステージ数
- 2)ステージにおける特徴量数
- 3)各ステージの閾値

が検出率と処理速度のトレードオフに作用する。各ステージの特徴量数は、誤検出率、検出もれ率がある程度得られるまで特徴量を追加して決定する。

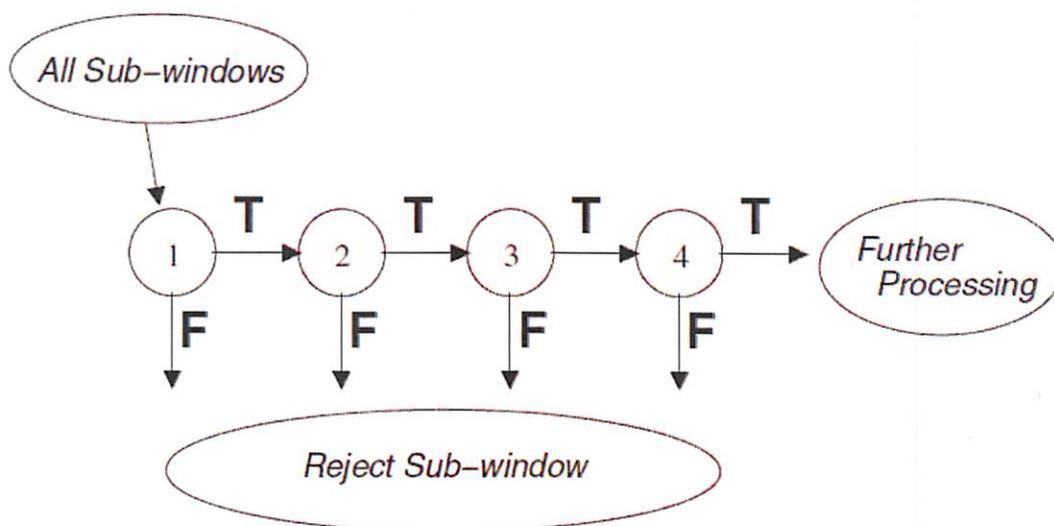


図 4.7 カスケード識別器の例(文献[29]より引用)

4.5 ニューラルネットワーク

ニューラルネットワーク(Neural network 神経回路網)は、脳機能に見られるいくつかの特性を計算機上のシミュレーションによって表現することを目指した数学モデルである[31]。研究の源流は生体の脳のモデル化であるが、神経科学の知見の改定などにより次第に脳モデルとは乖離が著しくなり、生物学や神経科学との区別のため、人工ニューラルネットワーク(Artificial Neural Network: ANN 人工神経回路網)とも呼ばれる[32]。

シナプスの結合によりネットワークを形成した人工ニューロン(ノード)が、学習によってシナプスの結合強度を変化させ、問題解決能力を持つようなモデル全般を指す。狭義には誤差逆伝播法を用いた多層パーセプトロンを指す場合もある。

ニューラルネットワークは、教師信号(正解)の入力によって問題に最適化されていく教師あり学習と、教師信号を必要としない教師なし学習に分けられる。明確な解答が用意される場合には教師あり学習が、データ・クラスタリングには教師なし学習が用いられる。結果としていずれも次元削減されるため、画像や統計など多次元量のデータでかつ線形分離不可能な問題に対して、比較的小さい計算量で良好な解を得られることが多い。このことから、パターン認識やデータマイニングをはじめ、さまざまな分野において応用されている。

4.5.1 基本構造

脳の中には多数のニューロン(神経細胞)が存在している。各ニューロンは、多数の他のニューロンから信号を受け取り、また、他の多数のニューロンへ信号を受け渡している。脳は、この信号の流れにより、様々な情報処理を行っている[33]。

この仕組みをコンピュータ内に実現しようとしたものがニューラルネットワークである。図(4.8)は、ニューラルネットワークの典型的な構造を示したものである。これらの図(4.8)において、円や四角は一つのニューロンに対応し、また、矢印が付いた枝は信号の流れを表している。

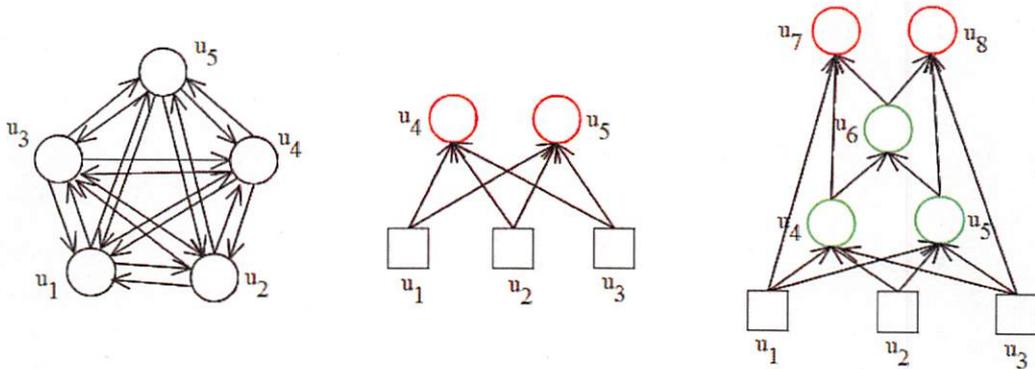


図 4.8 ニューラルネットワークのニューロンの例(文献[33]より引用)

左の図は、相互結合ネットワークを表し、また、中央と右の図は、階層的なネットワークを表している。四角で書かれたユニットは、入力ユニットと呼ばれ、外部からの信号を受け取る。また、赤いユニットは、出力ユニットと呼ばれ、外部へ信号を出力する。なお、相互結合ネットワークの場合は、すべてのユニットが、入力ユニットと出力ユニットを兼ねるケースが多く見受けられる。

中央及び右図のように、多層構造になっている場合、入力ユニットから構成されている層を入力層、また、出力ユニットから構成されている層を出力層と呼ぶ。右図の中間にあるユニット(緑色のユニット)は、隠れユニットと呼ばれ、その層は、隠れ層(中間層)と呼ばれる。

4.5.2 ニューロンのモデル化について

情報処理システムとしてのニューラルネットワークを設計するためには、生体ニューロンの動作を、数式で表現しなければならない。このように、自然界の事象を数式によって表現することをモデル化と呼ぶ。ところで、生体ニューロンの動作は、さらに細かく見れば、実際には極めて複雑になっている。したがって、生体ニューロンの動作の全てを忠実にモデル化することは不可能である。そこで、ニューラルネットワークを構成する一つの要素として、生体のニューロンの動作の本質的な特徴を失うことなく、なるべく簡単な数式でモデル化する必要がある。このような、ニューロンのモデルは、1943年に、マッカロックとピッツ(McCullochとPitts)によって最初に提案された。ニューロンモデルを図(4.9)に示した。

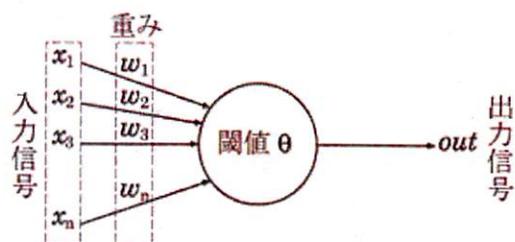


図 4.9 ニューロンのモデル(文献[33]より引用)

このニューロンモデルは、式(4.4)(4.5)に示すような数式に基づいて動作を行う。

$$\text{net} = w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (4.4)$$

$$\text{out} = f(\text{net} - \theta) \quad (4.5)$$

x_1, x_2, \dots, x_n は、他の n 個のニューロンからの入力信号を表し、 out はニューロンの出力信号を表す。 n 個の入力信号 x_1, x_2, \dots, x_n に対し、1 個の出力信号 out が決定される。また、 w_1, w_2, \dots, w_n は、シナプスの結合効率を表す量で、結合加重、あるいは、重みと呼ばれる。 net は生体のニューロンの膜電位に相当する量で、式(4.4)に示すように、入力信号と、それに対応する重みの積を n 個全て足し合わせるによって計算される。ここで、重みの絶対値が大きいと、入力信号が増幅されて伝わるので net に与える影響も大きくなり、逆に、重みの絶対値が小さいと net に与える影響も小さくなるのがわかる。 θ は、ニューロンが興奮するかしないかの閾値を表す。そして、式(4.5)に示すように、 net から θ を引いた値を活性化関数 $f(x)$ に代入して得られる値がそのまま、ニューロンの出力 out となる。マッカロックとピッツによって最初に提案されたときには、活性化関数は図 4.10 に示すような、ステップ関数が用いられている。

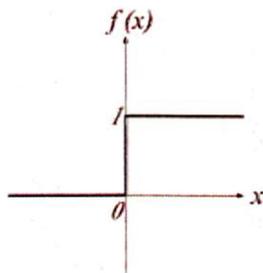


図 4.10 ステップ関数(文献[33]より引用)

4.5.2 ニューロンの学習について

ニューラルネットワークを作成しただけでは、入力ユニットにデータを与えても、出力ユニットから出る値が適切なものである可能性はほとんどない。何らかの方法で、ユニット間を繋ぐ枝の重み(バイアスも含む)を調整し、希望する出力が得られるようにしてやる必要がある。このことを行うのが、学習である。

学習方法には、大きく分けて二つの方法がある。それは、教師付き学習と教師無し学習である。教師付き学習は、幾つかの学習例と各学習例に対する目標出力を与え、目標出力と実際の出力が一致するように重みを調整する方法である。また、教師無し学習では、学習例からコンピュータ自身が何らかの基準に基づき重みを調整する。

また、この重みの調整には、学習例から1回の計算だけで行う方法と、学習例を繰り返し入力し、徐々に調整していく方法とがある。

4.5.3 誤差逆伝播法[33]

誤差逆伝播法に必要な重みの更新式は前節で求めた。それを使って、ここでは誤差逆伝播法による学習の流れをまとめてみる。

- 1) 初期値として、全ての重みを乱数によって-0.1~0.1程度の範囲の小さな値に設定する。さらに、学習率 η ($0 < \eta \leq 1$) を設定する。
- 2) 入力信号 x_{pi} ($1 \leq i \leq L_1$) をネットワークに入力する。
- 3) 入力層から出力層に向けて、各ニューロンの出力を計算する。
- 4) 出力層の出力 out_j^N と教師信号 y_{pj} の誤差から、 δ_j^N を計算する。具体的には、式(4.6)を計算する。

$$\delta_j^N = -(y_{pj} - out_j^N) out_j^N (1 - out_j^N) \quad (4.6)$$

- 5) 先のステップで求めた、 δ_j^N を使用し、中間層の誤差信号 δ_j^n ($n < N$) を計算する。具体的には、式(4.7)を計算する。

$$\delta_j^n = \left\{ \sum_{k=1}^{L_{n+1}} \delta_j^{n+1} w_{k,j}^{n+1,n} \right\} out_j^n (1 - out_j^n) \quad (4.7)$$

6) ステップ 4, ステップ 5 で求めた δ_j^n を使用し, 式(4.8)のように重みを更新する.

$$\Delta w_{ji}^{n,n-1} = -\varphi \delta_j^n out_j^{n-1} \quad (4.8)$$

全パターンに対する 2 乗誤差 E が十分に小さくなったと判断したら, 学習は終了する.
そうでなければ, 全パターンについてステップ 2 からステップ 6 までの操作を繰り返す.

以上が誤差逆伝播法の流れである.

4.6 おわりに

本章では, ターゲット検出を目的とした統計的学習手法について述べた. SVM, AdaBoost 及びニューラルネットワークについて詳しく説明しながら, 各アルゴリズムはマンガ画像との相性を検討した. 第 5 章では, 実験の詳細について述べる.

第 5 章

評価実験

5.1 はじめに

本章では、第 3 章で述べた特徴量と第 4 章で述べた統計的学習手法を用いて、マンガ画像におけるキャラクターの顔検出実験を行った結果を示し、その検出性能について検討する。

5.2 実験内容

本論文での実験はマンガ画像からキャラクターの顔の自動検出を対象とする。本実験の詳細を以下に示す。

5.2.1 実験目的

本実験では、マンガ画像から顔を検出する場合、HAAR-LIKE 特徴、HOG 特徴及び AdaBoost 学習アルゴリズムについての性能評価を目的とする。

5.2.2 実験環境

本実験では、WINDOWS 7 OS で OpenCV を用いて、実験を行った。OpenCV(正式名称: Open Source Computer Vision Library)は、オープンソースのコンピュータ・ビジョン・ライブラリである。コンピュータで画像や動画を処理するのに必要な、さまざま

機能が実装されており, BSD ライセンスで配布されていることから, 学術用途だけでなく商用目的でも利用できる. 加えて, マルチプラットフォーム対応されているため, 幅広い場面で利用されていることが特徴である[35].

5.2.3 実験データ

本実験では, 五つの作品に登場する主要なキャラクター5人 A, B, C, D, E について, 検出実験を行った. キャラクターの顔をポジティブ画像, 顔以外の部分をネガティブ画像とし, この五つ作品から選ばれた画像群に対して, 検出実験を行った. 学習に使用した画像枚数を表(5.1)に示す. ポジティブ学習画像とネガティブ学習画像の両方とも手動で切り出したものである. 切り出した学習画像を 64x64 に正規化してから学習させる. 本実験で使用するキャラクターの顔画像はすべて顔の正面画像である. 図(5.1)(5.2)(5.3)(5.4)はマンガ画像が含まれるエレメントの例, そして, 学習画像の例及び未知画像の例を示す.

表 5.1 本実験のデータセット

キャラクター	学習画像		未知画像	
	ポジティブ	ネガティブ	画像の数	顔の数
A	200	2000	50	96
B	200		50	91
C	200		50	93
D	200		50	101
E	200		50	95



図 5.1 マンガ画像が含まれるエレメントの例(文献[39]より引用)



図 5.2 一部ポジティブ学習画像の例(文献[39]より引用)

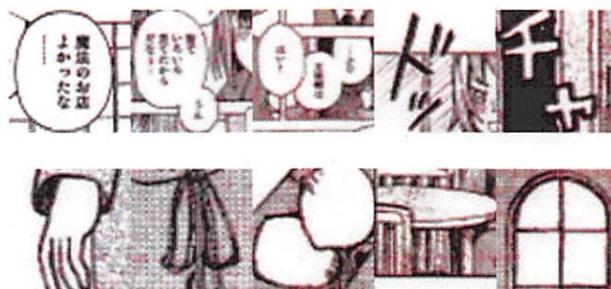


図 5.3 一部ネガティブ画像の例(文献[39]より引用)

5.2.4 評価標準

式(5.1)(5.2)(5.3)で各検出手法の性能を評価する。P, R, 及び F はそれぞれ適合率, 再現率, F 値を表す。

$$P = \frac{\text{検出された正しい数}}{\text{検出されたすべての数}} \quad (5.1)$$

$$R = \frac{\text{検出された正しい数}}{\text{検出されるべき数}} \quad (5.2)$$

$$F = 2 * (P * R) / (P + R) \quad (5.3)$$

5.3 実験手法

本実験では、マンガ画像からキャラクターの顔検出において、局所輝度情報を捉える HAAR-LIKE 特徴量とエッジ情報を捉える HOG 特徴量と LBP 特徴量, AdaBoost 学習アルゴリズムを組み合わせ、性能評価を行う[34][35]。まず、用意した学習画像を入力し、AdaBoost 学習アルゴリズムでキャラクターに別々学習させ、500 回繰り返し学習させた画像より各特徴量を算出し、データベースを構築する。次に、各キャラクターを切り出された作品を選んだ未知画像を入力し、検出実験をキャラクター毎に行う。各キャラクターで得た結果の平均値を求めて、HAAR-LIKE, HOG 及び LBP の性能を評価することになる[36][37]。

5.4 実験結果

マンガ画像の五つの作品合計 250 枚に含まれる未知画像を用いて顔検出実験を行った。実験結果を表(5.2)(5.3)(5.4)(5.5)に示す。

表 5.2 HAAR-LIKE 特徴量を用いて検出結果

キャラクター	未知画像	顔の数	P	R	F
A	50	96	0.37	0.74	0.49
B	50	91	0.41	0.78	0.54
C	50	93	0.39	0.71	0.50
D	50	101	0.45	0.78	0.57
E	50	95	0.29	0.63	0.40

表 5. 3 HOG 特徴量を用いて検出結果

キャラクタ	未知画像	顔の数	P	R	F
A	50	96	0.66	0.89	0.76
B	50	91	0.71	0.89	0.79
C	50	93	0.68	0.86	0.76
D	50	101	0.74	0.92	0.82
E	50	95	0.61	0.79	0.69

表 5. 4 LBP 特徴量を用いて検出結果

キャラクタ	未知画像	顔の数	P	R	F
A	50	96	0.69	0.83	0.75
B	50	91	0.64	0.89	0.74
C	50	93	0.73	0.85	0.79
D	50	101	0.71	0.90	0.79
E	50	95	0.68	0.83	0.75

表 5. 5 HAAR-LIKE, HOG 及び LBP の検出結果対比

特徴量	未知画像	P	R	F
HAAR-LIKE	250	0.38	0.73	0.50
HOG	250	0.68	0.87	0.76
LBP	250	0.69	0.86	0.77

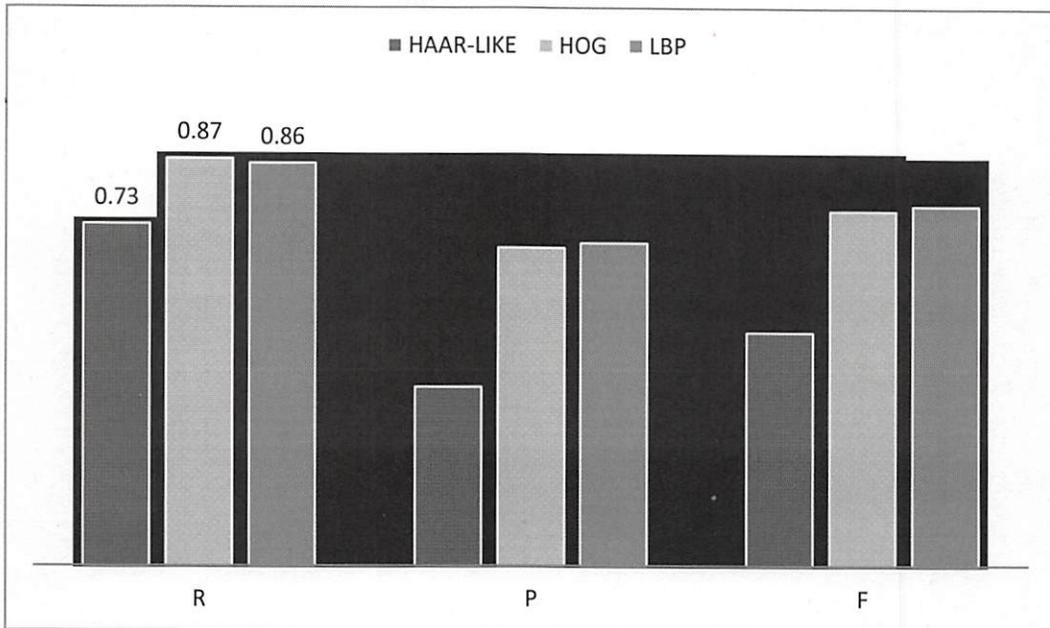


図 5.5 HAAR-LIKE, HOG 及び LBP 特徴の検出結果対比

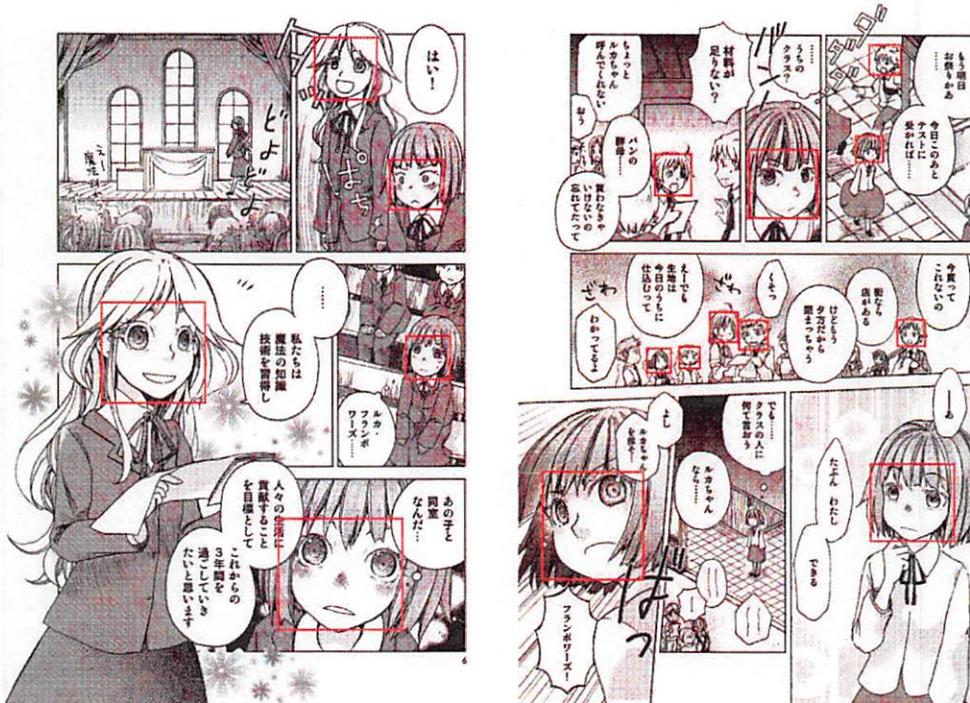


図 5.6 キャラクターの顔の正解領域の例(文献[39]より引用)

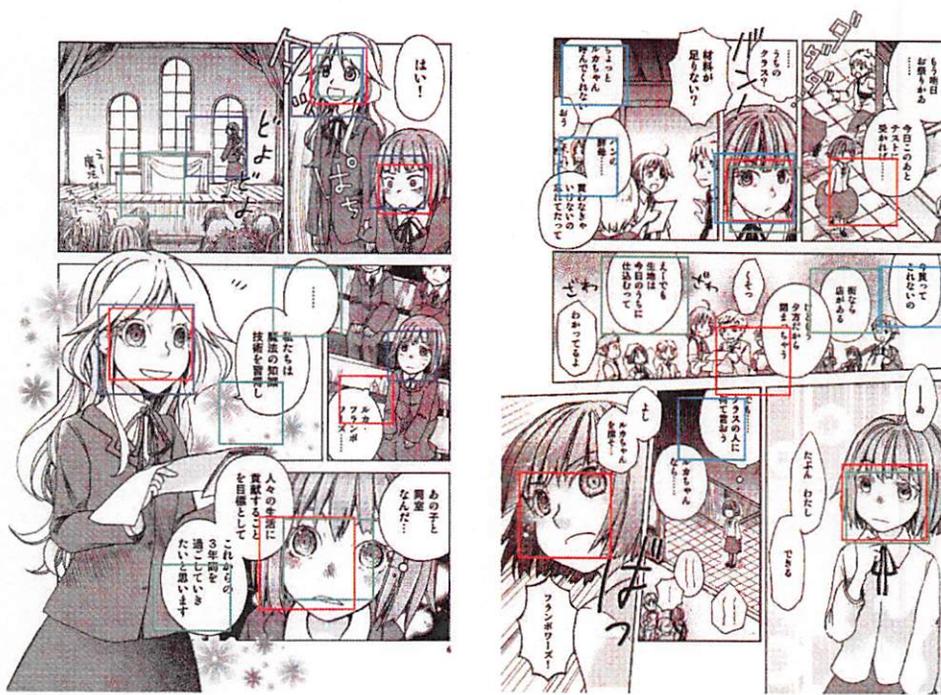


図 5.7 一部検出結果の例(文献[39]より引用)

5.5 考察

図(5.6)(5.7)はキャラクターの顔の正解領域の例及び一部検出結果の例を示した。検出結果より、HAAR-LIKE 特徴が HOG 特徴及び LBP 特徴より未知画像に対する検出結果は良くないことが分かる。その上で、未知画像に対する識別結果において、適合率である P 値が特に差があることが分かる。このことから、HAAR-LIKE 特徴量はマンガ画像に対して検出性能が低いと考えられる。一方で、HAAR-LIKE 特徴を利用した場合も、再現率である R 値が 0.73 となっており、マンガキャラクターの顔の特徴をうまく捉えられる場合もあると考えられる。

図(5.5)のように、評価尺度であり F 値は、HOG 特徴と LBP 特徴の場合には 0.76 と 0.77 であり、マンガ画像に対しては、おおよそ同じレベルの特徴把握能力を示した。検出結果全体から見ると、F 値が 0.5 である HAAR-LIKE 特徴より、HOG 特徴と LBP 特徴の方がマンガ画像からの顔検出能力が高いことが確認できた。

マンガは白黒二色で構成される 2 値画像である。主な要素は、点、線で描かれるキャラクター、背景とテキストである。現在、一般画像に対する顔検出手法のうち、広く用いられているのは Viola と Jones により提案された HAAR-LIKE 特徴量である。この特徴量は

様々な特徴パターンを用い、顔の各部分の輝度差を捉え、キャラクターの顔を検出する。一方でマンガキャラクターの顔は線により描かれるため、平均輝度に着目する HAAR-LIKE 特徴量では、良好な検出結果を期待することが困難だと考えられる。

また、本実験の学習画像の数はあまり多くないため、検出結果への影響が多少なりともあることがわかった。ただし、学習用画像数を増加させた場合でも検出成功率は一定程度で収束していることが確認できたため、実験結果はある程度有効であると考えられる。

5.6 おわりに

本章では、第 3 章で述べた特徴量と第 4 章で述べた統計的学習手法を用いて、マンガ画像におけるキャラクターの顔検出比較実験を行った結果を示し、その検出性能について検討した。

比較実験の結果により、HAAR-LIKE 特徴が HOG 特徴及び LBP 特徴より未知画像に対する検出結果は良くないことが分かった。また、未知画像に対する識別結果において、適合率である P 値が特に差があることが分かった。総合的に見て、HAAR-LIKE 特徴量はマンガ画像の顔検出性能が十分ではないと考えられる。

第 6 章では、本研究の研究目的、研究手法をまとめて述べる。さらに、今後の課題を検討する。

第 6 章

総括

6.1 まとめ

本研究では、マンガ画像からキャラクターの顔検出において、局所輝度情報を捉える HAAR-LIKE 特徴量とエッジ情報を捉える HOG 特徴量と LBP 特徴量, AdaBoost 学習アルゴリズムを組み合わせて、性能評価を行った。

第 1 章では、本研究の目的と必要性について述べた。

第 2 章では、近年の電子コミックマーケットの動向や、その閲覧端末について述べた。はじめに電子コミックマーケットの市場規模の変動やコンテンツ配信ビジネスの形態について述べ、次に電子コミックの将来展望について述べた。

第 3 章では、最近までに提案されてきている顔検出に適した特徴量について述べた。顔検出とは、画像中から顔の位置を検出する技術である。Haar-Like, SIFT, HOG 及び LBP 特徴を例として、近年で提案された顔検出向きの特徴量について述べた。そして、各特徴の算出方法についてそれぞれ紹介した。

第 4 章では、特徴量と組み合わせて統計的学習手法について述べた。ターゲット検出を目的とした統計的学習手法について述べた。SVM, AdaBoost 及びニューラルネットワークについて詳しく説明しながら、各アルゴリズムはマンガ画像との相性を検討した。

第 5 章では、HAAR-LIKE 特徴量, HOG 特徴量及び LBP 特徴量によりマンガ画像からキャラクターの顔検出対比実験を行った。検出結果より、HAAR-LIKE 特徴が HOG 特

徴及び LBP 特徴より未知画像に対する検出結果は良くないことが分かった。その上で、未知画像に対する識別結果において、適合率である P 値が特に差があることが分かった。このことから、HAAR-LIKE 特徴量はマンガ画像に対して検出性能が低いと考えられる。一方で、HAAR-LIKE 特徴を利用した場合も、再現率である R 値が 0.73 となっており、マンガキャラクターの顔の特徴をうまく捉えられる場合もあると考えられる。

第 6 章では、本論文の総括及び今後の課題について述べている。

6.2 今後の課題

今後の課題として以下の二つが挙げられる。

6.2.1 マンガ検出用学習サンプルデータベースの構築

統計的学習手法を用いて顔検出及び人検出の場合、本来に学習サンプルの数が多くほど多ければよいと言われ、一般画像で何万枚の学習サンプルが必要となる。しかし、従来のマンガ検出の場合、もともと顔サンプルのデータベースが存在せず膨大な学習サンプルが手に得られない状況にある。本研究では、学習サンプルの数がもっと多ければ、より正確な検出結果が得られたと考えられる。

6.2.2 新世代検出アルゴリズムの検討

本論文では比較的シンプルな特徴量を用いて、マンガ画像からの顔検出実験を行った。例えば、HOG 特徴、HAAR-LIKE 特徴、LBP 特徴などがこれに相当する。今後は、より複雑な次世代アルゴリズムについて検討する必要があると考える。例えば、deformable part model, ニューラルネットワークなどがこれに相当する。

謝辞

本研究を行うあたり、ご指導頂いた早稲田大学大学院国際情報通信研究科、渡辺 裕教授に心より感謝申し上げます。

本研究の機会を与えて下さり、研究の進め方から文章の書き方まで丁寧かつ御指導くださった石井 大祐氏には心より感謝申し上げます。

貴重なご意見そしてプログラミング支援を頂いた、共に研究に励んだ修士 2 年の張 傑氏に心から感謝致します。

論文を修正した時、ご助言を下された柳澤 秀彰氏に心から感謝致します。

また、渡辺研究室の皆様、研究室の仲間にも深く感謝いたします。

皆様、本当にありがとうございました。

参考文献

- [1] マガジン航
http://www.dotbook.jp/magazine-k/2014/10/15/future_of_e-comics/
- [2] Dalal N., Triggs B.: "Histograms of Oriented Gradients for Human Detection." IEEE Conf. CVPR, vol.1, pp.886-893, 2005.
- [3] 石井, 渡辺: "マンガからの自動人物検出と識別に関する一検討", 画像電子学会論文誌, Vol.42, No.4, pp.457-465, July 2013
- [4] 石井, 山崎, 渡辺: "マンガ固有の特徴を利用したマンガ登場人物識別に関する一検討", 情報処理学会AVM研究会研究報告, Vol.2013-AVM80, No.1, pp.1-4, Feb. 2013
- [5] デジタルコミック:
<http://ja.wikipedia.org/wiki/%E3%83%87%E3%82%B8%E3%82%BF%E3%83%AB%E3%82%B3%E3%83%9F%E3%83%83%E3%82%AF>
- [6] 植田尚子, 奥村友紀, 中山七彩: "マンガ産業の新たな可能性 "ISFJ 政策フォーラム, Dec. 2010
- [7] 高木ら, "電子書籍ビジネス調査報告書 2010," 三橋昭和, インプレス R&D, 東京, 2010.
- [8] 高木ら, "電子書籍ビジネス調査報告書 2013," 三橋昭和, インプレス R&D, 東京, 2014.
- [9] P. Viola and M. Jones: "Robust Real-Time Face Detection," Trans. IJCV, Vol.57, No.2, pp.137-154, 2004
- [10] 山内 悠嗣, 山下 隆義, 藤吉 弘亘: "[サーベイ論文]統計的学習手法による人検出", 電子情報通信学会 パターン認識・メディア研究会(PRMU) 技術報告, Vol. 112, No. 197, pp. 113-126, 2012.
- [11] R. Lienhart, and J. Maydt: An Extended Set of Haar-like Features for Rapid Object Detection, Proceedings of the IEEE International Conference on Image Processing, pp.900-903, 2002.
- [12] SIFT 特徴

http://en.wikipedia.org/wiki/Scale-invariant_feature_transform

[13] C. Harris and M. Stephens, "A combined corner and edge detector", Proc. of Fourth Alvey Vision Conference, pp. 147-151.

[14] 高木雅成, 藤吉弘亘, "SIFT 特徴量を用いた交通道路標識認識", 第 13 回画像センシングシンポジウム SSII07, LD2-06, 2007.

[15] 藤吉 弘亘, "Gradient ベースの特徴抽出", 情報処理学会 研究報告 CVIM 160, pp. 211-224, September, 2007.

[16] 石井, 渡辺: "マンガ画像からの自動キャラクター位置検出に関する検討", 情報処理学会 AVM 研究会研究報告, Vol.2012-AVM76, No.2, pp.1-5, Feb. 2012

[17] Papageorgiou, C. and Poggio, T. 'A trainable system for object detection', International Journal of Computer Vision, 38(1), 15-33. 2000

[18] D. Ishii, and H. Watanabe, "A Study on Frame Position Detection of Digitized Comics Images," Workshop Picture Coding Symposium, WP3-17, pp.124-125, Dec.2010.

[19] 藤吉研究室: "HOG 特徴量と Boosting を用いた人検出"
http://www.vision.cs.chubu.ac.jp/joint_hog/pdf/HOG+Boosting_LN.pdf

[20] LBP 特徴: compsci.world.coocan.jp/OUJ/2012PR/pr_15_a.pdf

[21] Y. D. Mu et al: "Discriminative Local Binary Patterns for Human Detection in Personal Album", CVPR, 2008.

[22] H. Yanagisawa, D. Ishii, H. Watanabe: "Face detection for comic images with deformable part model," The 4th International Workshop on Image Electronics and Visual Computing 2014 (IEVC2014), 4A-1, Oct. 2014

[23] C.W. Hsu and C.J. Lin: "A Comparison of Methods for Multi-class Support Vector Machines", IEEE Transactions on Neural Networks, Vol.13, pp.415-425, 2002.

[24] サポートベクターマシン:

<http://arx.ee.utsunomiya-u.ac.jp/research/svm/index.html>

- [25] C. C. Chang and C. J. Lin: "LIBSVM: a Library for support vector machines", ACM Transactions on Intelligent Systems and Technology, 2:27:1-27:27, 2011.
- [26] E. Osuna, R. Freund, and F. Girosi: "Training support vector machines: an application to face detection." Proc. of CVPR, pages 130–136, 1997.
- [27] Yoav Freund, Robert E. Schapire: "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting", 1995
- [28] AdaBoost: <http://ja.wikipedia.org/wiki/AdaBoost>
- [29] 藤吉研究室: "一般物体認識のための局所特徴量、SIFT と HOG"
<http://www.vision.cs.chubu.ac.jp/SIFT/>
- [30] C.-C. R. Wang, J.-J. J. Lien.: "AdaBoost Learning for Human Detection Based on Histograms of Oriented Gradients." ACCV2007, Part I , pp.885-895, 2007
- [31] ニューラルネットワーク:
<http://ja.wikipedia.org/wiki/%E3%83%8B%E3%83%A5%E3%83%BC%E3%83%A9%E3%83%AB%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF>
- [32] 村上研究室: <http://ipr20.cs.ehime-u.ac.jp/column/neural/>
- [33] 菅沼研究室:
http://www.sist.ac.jp/~suganuma/kougi/other_lecture/SE/net/net.htm
- [34] C.M.ピショップ: パターン認識と機械学習(上), p242-p248
- [35] OpenCV.jp: <http://opencv.jp/>
- [36] 奈良先端科学技術大学院大学: "OpenCV プログラミングブック," 2007
- [37] AdaBoost の実装
<http://mlab.cb.k.u-tokyo.ac.jp/~moris/lecture/cb-exp/AdaBoost.pdf>
- [38] 陳, 柳澤, 張, 石井, 渡辺: "マンガにおける HOG+AdaBoost による顔画像検出の性能評価", 映像情報メディア学会年次大会 17-4, Sep. 2014
- [39] 木野陽: ベリーベリークリームショコラ ふたつのベリー, 2010.

目次

第二章

図 2.1 現在スマートフォンで人気なマンガの閲覧アプリケーション図.....	5
図 2.2 電子書籍市場の割合と市場規模推移(文献[8]より引用)	6
図 2.3 電子コミックと紙コミックの市場規模対比(文献[8]より引用)	7
図 2.4 スマートフォン利用率の推移(文献[8]より引用)	7
図 2.5 電子書籍専用端末所有率の推移(文献[8]より引用).....	8

第三章

図 3.1 人間の顔のパターン(文献[10]より引用)	10
図 3.2 顔の明暗差分析図(文献[10]より引用)	11
図 3.3 HAAR-LIKE 特徴パターンの例	11
図 3.4 任意矩形領域の画素積分の計算式	12
図 3.5 マンガ画像例と漫画画像上の SIFT 特徴量(文献[16]より引用).....	14
図 3.6 1 セルの例(文献[19]より引用).....	15
図 3.7 HOG 特徴のヒストグラム(文献[19]より引用)	15
図 3.8 1 ブロックの例(文献[19]より引用)	16
図 3.9 LBP 特徴の算出手法(文献[21]より引用)	17

第四章

図 4.1 一般的な統計的学習手法による物体検出の流れ(文献[15]より引用).....	19
図 4.2 識別の例(文献[24]より引用)	20
図 4.3 SVM によるマージン最大化処理(文献[24]より引用)	21
図 4.4 線形と非線形 SVM の例(文献[24]より引用)	22
図 4.5 弱識別器を選択するルール(文献[39]より引用)	24
図 4.6 エラー率を算出の例(文献[29]より引用)	24
図 4.7 カスケード識別器の例(文献[29]より引用).....	25
図 4.8 ニューラルネットワークのニューロンの例(文献[33]より引用).....	27
図 4.9 ニューロンのモデル(文献[33]より引用).....	28

図 4. 10 ステップ関数(文献[33]より引用)	28
----------------------------------	----

第五章

図 5. 1 マンガ画像が含まるエレメントの例(文献[39]より引用)	33
図 5. 2 一部ポジティブ学習画像の例(文献[39]より引用)	33
図 5. 3 一部ネガティブ画像の例(文献[39]より引用)	33
図 5. 4 一部未知画像の例(文献[39]より引用)	34
図 5. 5 HAAR-LIKE, HOG 及び LBP 特徴の検出結果対比	37
図 5. 6 キャラクターの顔の正解領域の例(文献[39]より引用)	37
図 5. 7 一部検出結果の例(文献[39]より引用)	38

表目次

第四章

表 4. 1 HOG 及び SVM による顔検出実験	22
----------------------------------	----

第五章

表 5. 1 本実験のデータセット	32
表 5. 2 HAAR-LIKE 特徴量を用いて検出結果	35
表 5. 3 HOG 特徴量を用いて検出結果	36
表 5. 4 LBP 特徴量を用いて検出結果	36
表 5. 5 HAAR-LIKE, HOG 及び LBP の検出結果対比	36

研究業績

	題名	発表年月	発表掲載誌	連名者
(1)	“A Note on Face Detection of Comic Image with Different Background	2014 年 3 月	2014 IEICE General Conference	M. Chen, H. Yanagisawa, D. Ishii, H. Watanabe
(2)	マンガにおける HOG+AdaBoost による顔画像検出の性能評価	2014 年 9 月	映像情報メディア学会年次大会	陳 明 柳澤 秀彰 張 傑 石井 大祐 渡辺 裕
(3)	マンガ画像からの顔検出におけるパーツ特徴量の一検討	2014 年 9 月	映像情報メディア学会年次大会	柳澤 秀彰 石井 大祐 陳 明 渡辺 裕