

Summary of Bachelor's Thesis
2014年3月修了卒業論文概要書

Name 氏 名	柳澤秀彰	ID number 学籍番号	1W100499-7
題 目 Title (日本語の場合は英文題目も記入)	Supervisor 指導教員 渡辺 祐		印
日本語 Japanese Deformable Part Modelを用いたコミック画像からの顔検出			
英語 English Face Detection for Comic Images with Deformable Part Model			

概 要 Summary

電子コミックの製作作業を効率化するために、コマ割り、吹き出し、キャラクター位置といったメタデータを自動的に抽出する技術についての研究が広がっている。現時点において、コマ割りと吹き出しの抽出に関しては精度の高い手法が提案されている。一方、キャラクターの検出については、HOG特量を用いることで、コミック画像内に描かれたキャラクターの顔の位置の検出が行なえることが示されている。しかし、顔以外の部分の誤検出が多いという問題がある。

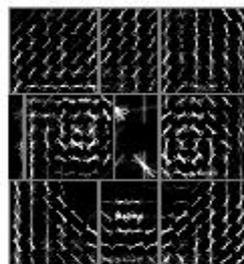
そこで本論文では、HOG特徴量に加えて顔パーツの配置特徴を利用することで、顔検出の精度向上を図った。

パーツ配置を利用した手法として、Felzenszalbらによって提案された物体検出手法であるDeformable Part Model を使用する。Deformable Part Modelは現在までに、人や車、動物といったさまざまな自然物体に対して検出実験が行なわれているが、コミック画像のような線画で表現された物体に対しても有効であるかは未知である。本論文では、Deformable Part Model をコミック画像に適用した場合の性能評価を行なった。

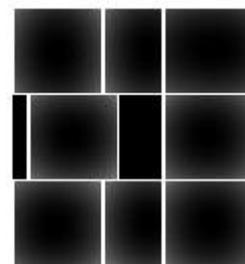
実験の結果、80%以上の顔領域に対して検出が可能であった。また、検出した領域の正解率は最大で97.3%であった。これより、Deformable Part Modelがコミック画像に対しても物体検出が可能であることが分かった。また、コミック画像における物体の検出において、パーツ配置の特徴を利用することが効果的であることが分かった。



(a)ルート特徴



(b)パート特徴



(c)パートフィルタ配置

Deformable Part Modelによる顔領域の検出モデル

- 注意 1. 提出期限/ Deadline of Submission 2014年2月6日(木)17:00 /February 6, 2014 17:00
2. 原本1部、コピー1部の計2部提出 / Must submit one original and one copy

基幹理工学部情報理工学科

Bachelor's Thesis
卒業論文

Title
論文題目

Deformable Part Modelを用いたコミック画像からの顔検出

Face Detection for Comic Images with Deformable Part Model

Student ID 学籍番号	1W100499-7
Name 氏名	Hideaki Yanagisawa 柳澤 秀彰

Supervisor 指導教員	Hiroshi Watanabe 渡辺 裕	印
--------------------	--------------------------	---

2014年3月15日

概要

電子コミックの製作作業を効率化するために、コミック画像からコマ割り、吹き出し、キャラクター位置といったメタデータを自動的に抽出する技術についての研究が広がっている。現時点において、コマ割りと吹き出しの抽出については精度の高い手法が提案されている。一方、キャラクターの検出については、HOG 特量を用いることで、コミック画像内に描かれたキャラクターの顔の位置の検出が行なえることが示されている。しかし、顔以外の部分の誤検出が多いという問題がある。

そこで本論文では、HOG 特徴量に加えて顔パーツの配置特徴を利用することで、顔検出の精度向上を図った。

パーツ配置を利用した手法として、Felzenszalb らによって提案された物体検出手法である **Deformable Part Model** を使用する。**Deformable Part Model** は現在までに、人や車、動物といったさまざまな自然物体に対して検出実験が行なわれているが、コミック画像のような線画で表現された物体に対しても有効であるかは未知である。本論文では、**Deformable Part Model** をコミック画像に適用した場合の性能評価を行なった。

実験の結果、80%以上の顔領域に対して検出が可能であった。また、検出した領域の正解率は最大で 97.3%であった。これより、**Deformable Part Model** がコミック画像に対しても物体検出が可能であることが分かった。また、コミック画像における物体の検出において、パーツ配置の特徴を利用することが効果的であることが分かった。

目次

第1章	はじめに	3
1.1	研究の背景	3
1.2	本研究の目的	3
1.3	論文の構成	4
第2章	画像認識の基礎技術	5
2.1	HOG(Histograms of Oriented Gradients)	5
2.2	Deformable Part Model	5
2.2.1	検出モデル	5
2.2.2	HOG ピラミッド	6
2.2.3	フィルタ	7
2.2.4	可変パーツ	7
2.2.5	検出	7
2.2.6	学習	8
2.2.7	Latent-SVMs	9
2.2.8	ハードネガティブの抽出	9
2.2.9	学習の詳細	9
第3章	コミック画像のキャラクター顔画像検出	11
3.1	提案手法	11
3.2	実験の概要	12
3.3	実験条件	12
3.4	実験結果	14
第4章	考察	17
第5章	結論	19
	謝辞	20
	参考文献	21
	図一覧	22
	表一覧	23

第1章 はじめに

1.1 研究の背景

近年、タブレットやスマートフォンといったデジタル端末で読むことのできる電子書籍の発展が顕著である。2012年度の電子書籍市場規模は729億円と推計され、2011年度の629億円から100億円増加している^[1]。その中で電子コミックは電子書籍市場の中で大きな地位を担うコンテンツであり、2011年度に発表された電子書籍ビジネス調査報告書によると、電子コミックの売上高は514億円にのぼり、これは電子書籍市場全体の約81.7%を占めている^[2]。これより、今後も電子書籍市場の規模が拡大すると予想されるなかで電子コミックは重要な存在であるといえる。

そうした中で、電子コミックの新たな可能性を模索した研究が広がりつつある^[3]。電子コミックを紙媒体の従来のコミック作品と比較した際に挙げられるメリットとして、物理的な制約がないために、従来のコミックの枠にとられない表現や利用が可能であるという点がある。例としては、コミック内のキャラクター、台詞、シーンなどの情報を記録することで、読み手の母語に応じてコミック内の台詞やオノマトペといった文字データの言語を切り替えるといった機能や、コミックのタイトルや作者情報だけでなく、特定のキャラクターやシーンの情報を元にアーカイブよりコミック作品の検索や提示を行なうといった機能が提案されている。このようなユーザーセントリックな機能の提供が、今後の電子コミック市場において大きな価値となることが期待できる。前述の技術の実現化には、デジタル化されたコミック画像のアーカイブにおいて、キャラクター位置や、コマ割り、吹き出しといったコミックの内容に関するメタデータを抽出し、それらを重畳して元のデータと併せて保存する必要がある。しかし、現状でこれらのメタデータを抽出するためには、紙媒体からスキャンしてデジタル化した電子コミック画像から、手作業で切り出してデータを抽出する必要があるため、コストがかかるという問題がある。よって、より作業を効率化するために、自動でコミック画像よりメタデータを抽出する技術が必要とされている。

現在、コミック画像よりコマ割り情報を抽出する技術については、コミックの枠線を識別し、濃度勾配 (*intensity gradient*)の方向を利用してコマの分割線を同定する手法^{[4][5][6]}や、「コミックのコマは矩形であることが多い」という特徴を利用して、画像内から矩形領域を検出し、コマを特定する手法^[7]などが提案されており、いずれの手法でも80%を越える精度が報告されている。また、吹き出しを同定する技術については、ページ内の文字領域をAda Boostによって特定し、その領域をもとに吹き出し候補を検出、またSVMによって吹き出し形状の分類を行なう手法^[8]が提案されており、この手法によって86%の吹き出しの同定が行なえることが報告されている。一方で、コミックに登場するキャラクターを抽出する技術については、HOG特徴量 (*Histograms of Oriented Gradients*)^[9]を利用して画像内から顔候補となる領域を特定し、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うことで、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^{[10][11]}が提案されているが、キャラクターによる精度のばらつきが大きいいため、現状において安定的な検出ができていないという問題がある。

1.2 本研究の目的

本研究では、背景に述べた点を踏まえ、コミック画像におけるより高精度なキャラクター顔検出の実現を目的とする。コミック画像から顔検出を行なう技術については、過去にHOG特徴量とSVMを利用した手法が提案されている。この手法による検出結果の例を図1に示す。



図1 HOG 特徴量を用いた学習による検出結果^[11]. (赤い四角が顔と判断された箇所を示す. 左側が学習画像を含む既知の画像, 右側が未知の画像.)

図1に示す実験結果より, キャラクターの顔領域において正の検出結果を得ることができている. また, この研究ではキャラクターの顔領域全体を学習するより, 瞳領域を検出対象として学習した場合の方が, より精度の高い検出が行なえることが示されている. しかし, 瞳領域を検出した場合においても, 未だに偽の領域が誤って検出されている箇所が存在しており, 現状では安定した検出が行なえていない.

今回, より精度の高い検出を行なうための新たな手法として, パーツ配置の特徴を利用した物体検出モデルである **Deformable Part Model**^{[12][13]}を使用することを考え, コミック画像について当手法が有効であるか検証を行なった.

1.3 論文の構成

本論文では, 第2章で画像認識の基礎技術について述べ, 第3章でコミック画像のキャラクター顔画像の検出処理について述べ, 第4章で実験結果から考察を行なう. 最後に第5章で, 本論文の結論と今後の課題について述べる.

第 2 章 画像認識の基礎技術

この章では、本論文で使用する概念や用語の定義について説明を行なう。

2.1 HOG (Histograms of Oriented Gradients)

HOG^[9]は、人検出のため、Dalal らによって提案された特徴量である。局所領域における輝度の勾配方向をヒストグラム化した特徴ベクトルであり、照明変動の影響を受けにくく、局所的な幾何学変化に頑強であるといった性質を持つ。HOG 特徴量は以下の手順から算出される。

1. 画像の各ピクセルの輝度から勾配強度と勾配方向を求める。このとき、勾配方向は $0^{\circ} \sim 160^{\circ}$ までの範囲を 20° ずつ区切った 9 方向に離散化される。
2. 局所領域を 8×8 ピクセルから成るセルに分割する。
3. 各セル領域において、輝度の勾配方向ヒストグラムを作成し、9 次元のベクトルを得る。
4. 2×2 セルを一つのブロックとして、セルのベクトルを結合して正規化を行い、ブロックを表すベクトルを得る。最終的に得られる特徴ベクトルは、全てのブロックのベクトルから構成される多次元ベクトルとなる。

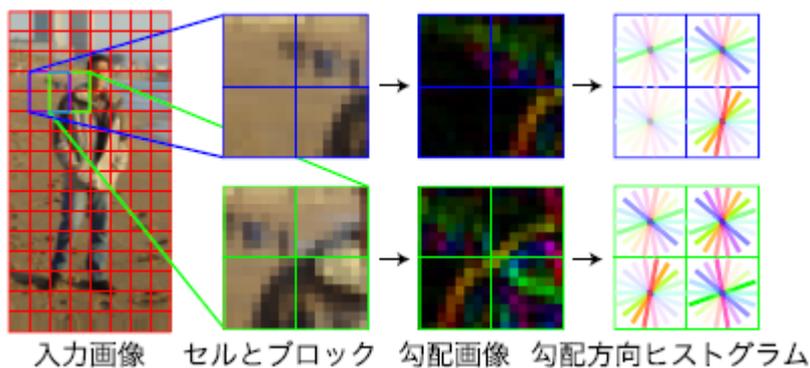


図 2 HOG 特徴量^[14]

2.2 Deformable Part Model

Deformable Part Model は、物体のモデルをパーツの集合として表現し、それぞれのパーツの妥当性およびそれらの相対位置関係で評価を行なう物体検出手法である。従来の方式では、検出対象のパーツの位置は固定されており、人や動物といった物体を検出する際に姿勢が大きく変化した場合に対応できないという問題があった。この方式では、検出対象のパーツの位置が可変であるため、画像によって姿勢が大きく変化する対象物に対しても高いパフォーマンスを示すことが可能である。この方式は 2006 年の PASCAL Person Detection Challenge において、最高のパフォーマンスについて従来手法の 2 倍の平均精度を成し遂げ、20 のカテゴリ中 10 で最高の結果を出している^[15]。

2.2.1 検出モデル

検出対象モデルは、検出対象全体を捉えるグローバルなルートフィルタと、検出対象が持ついくつかのパートフィルタから構成される。特徴量記述子としては HOG 特徴量を使用し、ルートフィルタにおいて対象全体の疎な特徴を、パートフィルタにおいてパーツごとの密な特徴を捉える。図 3 に Deformable Part Model を用いた人検出の例を示す。

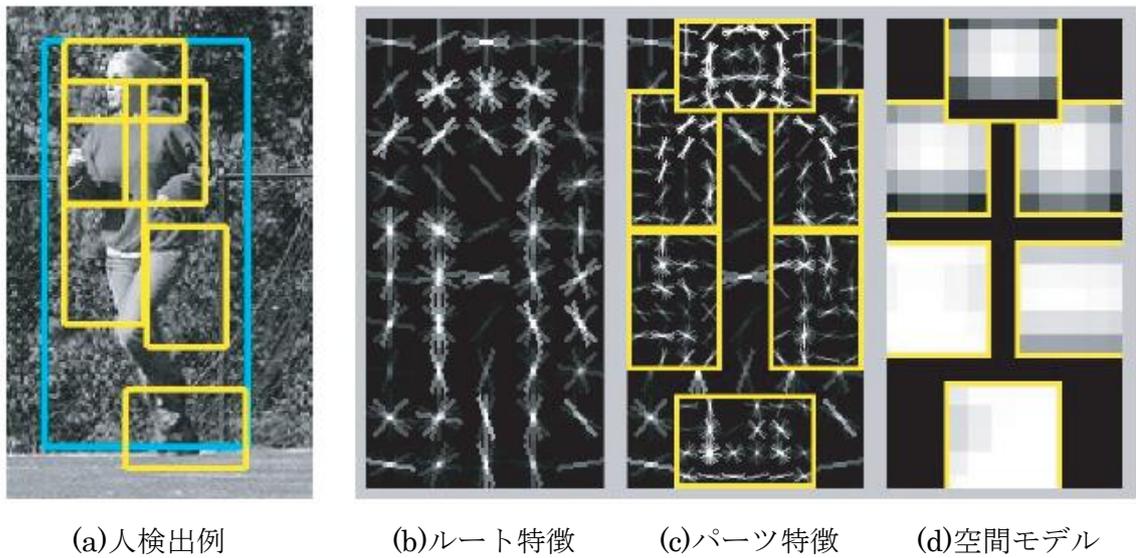


図3 Deformable Part Model の人検出モデル^[15]

2.2.2 HOG ピラミッド

Deformable Part Model では、ルートフィルタとパートフィルタのそれぞれのスケールに対して HOG 特徴量を適用する。対象となる画像のスケールを変化させて、解像度の異なる画像の集合であるイメージピラミッドを作成する。イメージピラミッドの各々のレベルの HOG 特徴を計算することによって、HOG 特徴量ピラミッドを求める。ここで、イメージピラミッドの上層で大域的に粗い HOG 特徴を捉え、下層で局所的に細かい HOG 特徴を捉える。

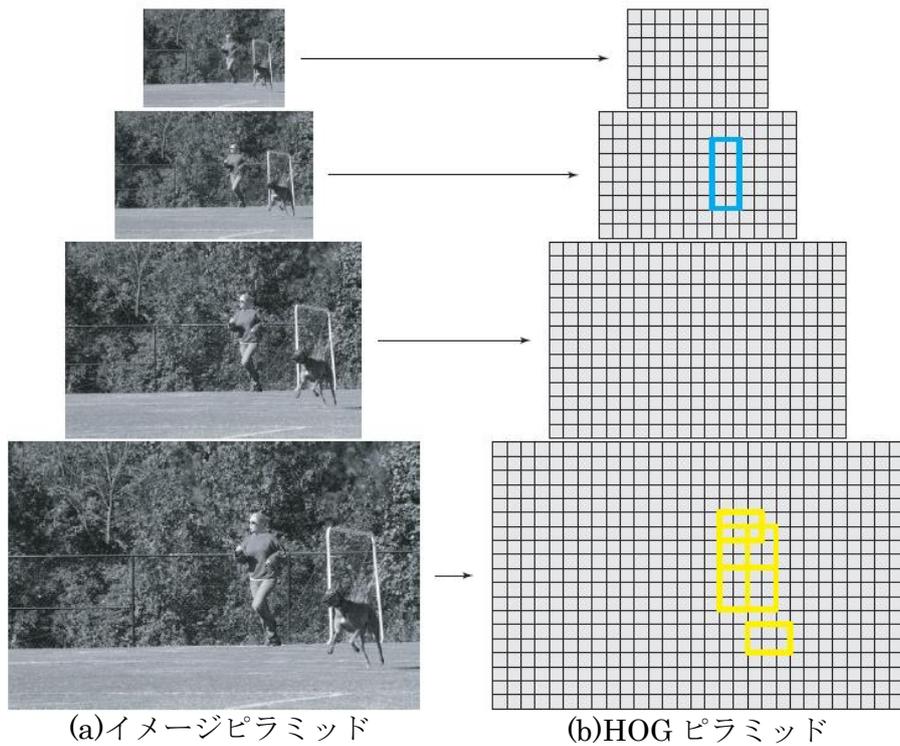


図4 イメージピラミッド^[15]

2.2.3 フィルタ

フィルタ F は、入力特徴に対する重みであり、 $w \times h \times 9 \times 4$ 個のベクトルで表現される。 w と h は、サブウィンドウの高さと横幅を表している。 HOG ピラミッドを H 、セルの位置を $p = (x, y, l)$ で表す。 l は、HOG ピラミッドのレベル (階層) を、 $\phi(H, p, w, h)$ は、取得する HOG 特徴の強さを示す。 検出ウィンドウにおけるフィルタ F のスコアは、重みを持ったベクトルと、特徴量の内積 $F \cdot \phi(H, p, w, h)$ で表される。

2.2.4 可変パーツ

ルートフィルタは検出ウィンドウと同等と定義し、パートフィルタでのセルのサイズは、ルートフィルタのレベルにおけるセルのサイズの半分となるようにする。 このように、ルートフィルタより物体の概形を見るよりは、パートフィルタより高い解像度でのパーツの特徴を見るほうが局所的であり、高い認識性能を得ることができると考えられる。

n 個のパーツでできた物体のモデルはルートフィルタ F_0 とパートモデル (P_1, \dots, P_n) とで表される。 このとき $P_i = (F_i, v_i, s_i, a_i, b_i)$ である。 ここで、 F_i は i 番目のパートフィルタ、 v_i はルートフィルタと i 番目のパートフィルタの中心座標の相対的な位置関係を示す 2 次元ベクトル、 s_i は i 番目のパートフィルタの中心点を定める際の許容範囲を定める際の許容範囲を表すボックスのサイズ、 a_i, b_i は、 i 番目のパートフィルタにおける 2 次元ベクトルによる係数を表す。

モデルの配置を潜在変数 z とし、 $z = (p_1, \dots, p_n)$ で与える。 配置 z のスコアは、各フィルタのスコアと、パーツとルートの位置関係により、(1)式で与えられる。

$$\text{score}(z) = \sum_{i=0}^n F_i \cdot \phi(H, p_i) - \sum_{i=1}^n a_i \cdot (\tilde{x}_i, \tilde{y}_i) + b_i \cdot (\tilde{x}_i^2, \tilde{y}_i^2) \quad (1)$$

$$(\tilde{x}_i, \tilde{y}_i) = ((x_i, y_i) - 2(x, y) + v_i) / s_i \quad (2)$$

(1)式において、第一項目は、フィルタの重みと、HOG 特徴ベクトルの内積をとったフィルタのスコアの合計を表しており、第二項目は、パートフィルタとルートフィルタの相対的な位置関係と距離を表している。

(2)式は、 i 番目のパートフィルタの中心座標 (x, y) と、ルートフィルタの中心座標 (x_i, y_i) 、 v_i 、 s_i を用いて配置を表す $(\tilde{x}_i, \tilde{y}_i)$ を算出する。 このとき、パートフィルタはルートフィルタの 2 倍の解像度を持っているので、距離関係を元に戻すために、 (x, y) に 2 を掛け合わせて計算する。 ここで \tilde{x}_i と \tilde{y}_i は共に -1 から 1 の値をとる。 また、式(1)での配置 z のスコアは、次式の β と $\psi(H, z)$ の内積 $\beta \cdot \psi(H, z)$ で表せる。

$$\beta = (F_0, \dots, F_n, a_1, b_1, \dots, a_n, b_n) \quad (3)$$

$$\psi(H, z) = (\phi(H, p_0), \phi(H, p_1), \dots, \phi(H, p_n), \tilde{x}_1, \tilde{y}_1, \tilde{x}_1^2, \tilde{y}_1^2, \dots, \tilde{x}_n, \tilde{y}_n, \tilde{x}_n^2, \tilde{y}_n^2) \quad (4)$$

2.2.5 検出

スライディングウィンドウによって各ルート位置におけるスコアを求め、スコアの値が高くなった位置に対象物体があるとして検出する。 このとき、ルート位置のスコアは式(1)で表されるスコアの最大値となる。

$$\text{score}(p_0) = \max_{p_1, \dots, p_n} \text{score}(z) \quad (5)$$

$$\text{score}(p_0) = \max_{p_1, \dots, p_n} \left(\sum_{i=0}^n F_i \cdot \phi(H, p_i) - \sum_{i=1}^n a_i \cdot (\tilde{x}_i, \tilde{y}_i) + b_i \cdot (\tilde{x}_i^2, \tilde{y}_i^2) \right) \quad (6)$$

また、各パーツは独立なので、それぞれについて最大値を求めることでルート位置のスコアを求めることができる。

$$\text{score}(p_o) = F_0 \cdot \phi(H, p_i) + \sum_{i=1}^n \max_{p_i} F_i \cdot \phi(H, p_i) - (a_i \cdot (\tilde{x}_i, \tilde{y}_i) + b_i \cdot (\tilde{x}_i^2, \tilde{y}_i^2)) \quad (7)$$

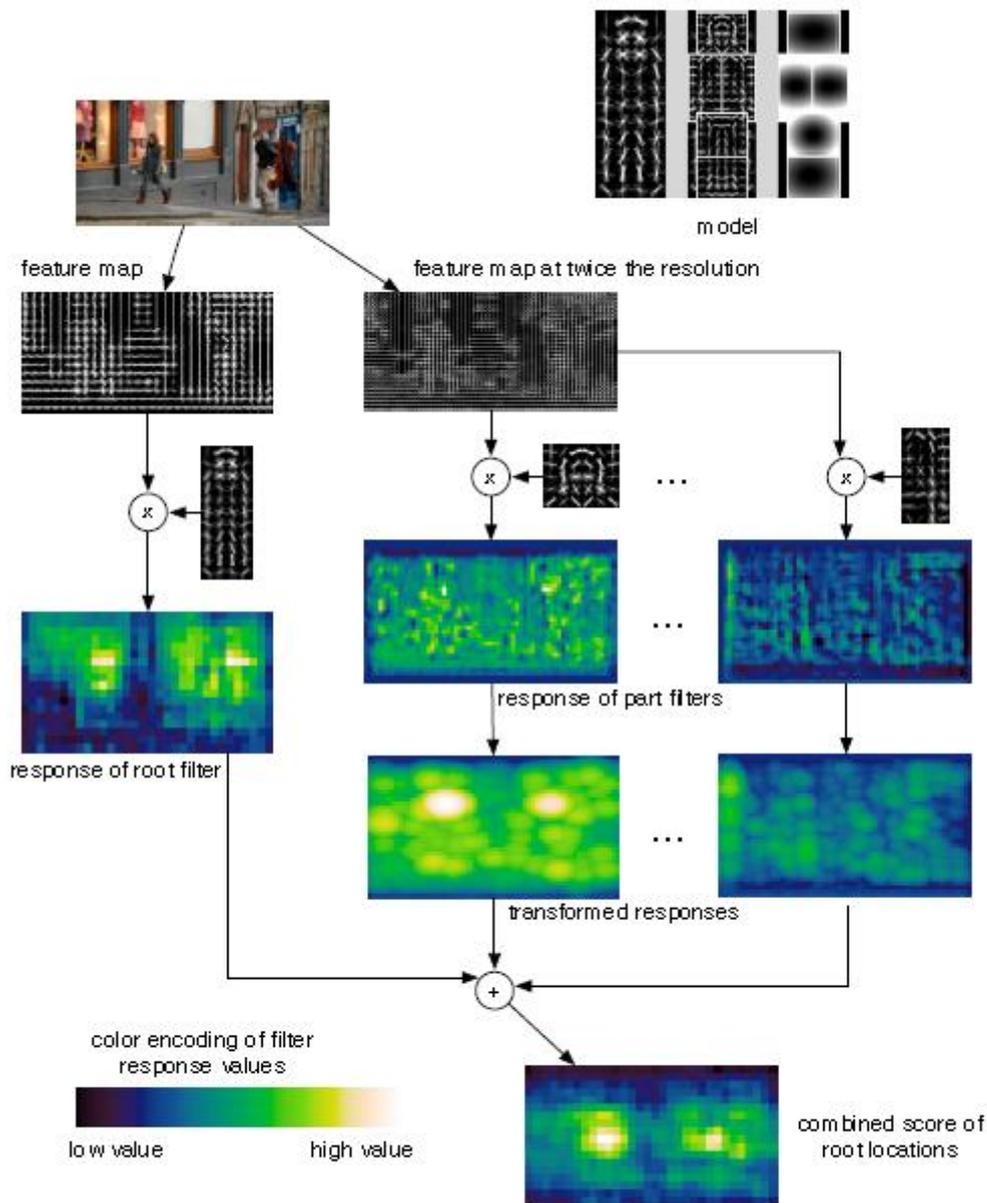


図5 Deformable Part Model のスコア算出の流れ^[13]

2.2.6 学習

学習には対象物体のバウンディングボックスを指定したポジティブサンプル画像と、検出対象ではないものだけが含まれているネガティブサンプル画像を合わせたデータセットを用いる。学習データセット $D = (\langle x_1, y_1 \rangle, \dots, \langle x_n, y_n \rangle)$ とし、 $y_i \in \{-1, 1\}$ はサンプル x_i に対するラベルとする。また x_i より、HOGピラミッドを $H(x_i)$ 、ルートフィルタと

パートフィルタの有効な配置の範囲を $Z(x_i)$ とする。この $Z(x_i)$ は、ポジティブサンプルにおいて、指定されているバウンディングボックスを元に決められ、最低でも 50%の領域が重なるように定義される。ポジティブサンプルより、パートの位置ごとと対象物体そのものの位置、および各フィルタの入力特徴に対する重みについて学習を行なう。

2.2.7 Latent SVMs

x のそれぞれのサンプルにおけるスコアは次式で表される。

$$f_{\beta}(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \quad (8)$$

ここで、 β はモデルのパラメータ、 z はモデルの配置を表す潜在変数である。

$\Phi(x, y) = \psi(H(x), z)$ と定義し、置き換えることができるので、式(8)では、式(3)(4)で示した配置のスコアの最大値をとることと等しくなる。よって、 β を学習することでポジティブサンプルの学習を行なうことができる。 β や z といった潜在的な変数を使い学習するアプローチを Latent SVM と呼ぶ。Latent SVM はまず β を固定し、 $f_{\beta}(x)$ を最大化する z を求める、次に z を固定して通常の SVM のアルゴリズムにより β の最適化を行なう。この操作を繰り返すことによって、最適な β の値を求める。

2.2.8 ハードネガティブの抽出

一般物体認識において、膨大な量の学習サンプルの大多数はネガティブサンプルとなる。一度に全てのネガティブサンプルを用いることは難しいため、ネガティブサンプルの中からより識別しづらいものをハードネガティブサンプルとして選び、ポジティブサンプルと、ハードネガティブサンプルから成る学習データを作成する。ハードネガティブサンプルの作成は D と β を用いて次式で表され、間違えて識別をした学習サンプルの集合となる。

$$M(\beta, D) = \{(x, y) \in D | y f_{\beta}(x) \leq 1\} \quad (9)$$

2.2.9 学習の詳細

Deformable Part Model の学習プロセスの詳細は以下のようにになる。

1. ルートフィルタの初期化

学習用サンプルで定められたバウンディングボックスからルートフィルタの寸法を自動決定する。初期の学習では、潜在変数を持たない通常の SVM を用いてルートフィルタ F_0 を学習する。ポジティブサンプルはデータセットで用意されているものを使用し、ネガティブサンプルはデガティブイメージの中からランダムに決定される。

2. ルートフィルタの更新

バウンディングボックスに重なるように式(1)の配置のスコアが最も高いルートフィルタを探し出して更新する。その後、 F_0 を再学習する。

3. パートフィルタの初期化

上記より学習されたルートフィルタから、ルートフィルタの 80%を占めるように 6 つのパートフィルタを採用する。一番初めのパートフィルタは、HOG 特徴量が最もポジティブな位置とし、六個のパートフィルタの位置を順に決定していく。初期の変化コストは、パートフィルタの係数 a_i, b_i の初期値 $a_i = (0, 0)$, $b_i = -(1, 1)$ より求める。

4. モデルの更新

Latent SVM により、新しいモデルを更新するため、学習データにモデルの配置 z を追加して、 $\langle x_i, z_i, y_i \rangle$ という形で再構築する。その後、バウンディングボックスに最低 50%重なるように画像より検出を行なう。この中で、バウンディングボックスの配置と最もよく一致しているものを採用する、これにより β も更新される。ハードネガティブサンプルは、対象物体ではないのに高いスコアを出してしまったものを選ぶ。これをファイルのサイズの限界まで 10 回の学習を繰り返し行なう。学習の過程において、ハードネガティブサンプルを記録しておき、メモリの限界の範囲内において、できるだけ多

くの新しいハードネガティブサンプルを追加していく.

第3章 コミック画像からの顔画像の検出

3.1 提案手法

一般的な画像解析技術における物体の検出処理は、画像中の輝度などから得られる情報を特徴量として記述し、統計的な解析と学習により判別器を構成して、画像の中に目標とする物体が存在するか検出を行なうというものである。コミック画像は自然画像と異なる信号的特徴を持ち、かつ人物等のオブジェクトが自然界に存在するものとは異なった表現で描かれているため、コミック画像を対象とした解析処理を実現するためには、コミック画像に対する処理内容に応じた特徴量を適切な形で利用する必要がある。コミック画像は主に以下の三つの組み合わせによって構成されている。

1. 白黒の2値による線画.
2. 塗り潰し領域やトーン等のテクスチャ領域.
3. 台詞などの文字領域.

これより、コミック画像は自然画像と比較して、画像中の信号として輝度の変化が大きいエッジ成分と輝度の変化が平坦な領域が多く含まれているといえる。よって、エッジ成分を多く含んでいるコミック画像の特徴を表現する記述子としては、局所領域の輝度勾配を利用する HOG 特徴量が適していると考えられる。

しかしながら、前述の通り、既存の HOG 特徴量を用いた手法では、コミック画像から安定した顔検出を行えていない。その理由としては、コミックに登場するキャラクターの顔は一般に線画で表現されているため、実画像の顔認識に比べて識別に利用できる特徴量が限られていることや、コミック特有の誇張表現のために顔や輪郭のぼらつきが大きいことなどが挙げられる。よって、キャラクターの顔領域を正確に捉えるためには、HOG 特徴量と併せて別の特徴を利用する必要があると考えられる。よって今回、画像の輝度変化以外の特徴を利用することで、キャラクターの検出精度を高めることを考えた。

登場シーンごとのキャラクターの顔部分の変化例を図 6 示す。これより、登場シーンごとにキャラクターの顔特徴が大きく変化することが示されているが、一方で目や口といった顔パーツの配置は多くのシーンにおいて普遍であることが分かる。今回、顔パーツの配置情報を比較することで、顔部分とそれ以外の領域との判別をより高い精度で行なうことができるのではないかと予測した。それを踏まえて、新たな手法として Felzenszwalb らが提案した物体検出手法である Deformable Part Model が適していると考えた。この手法は、対象物体の HOG 特徴量と、物体の持つパーツ配置の特徴を組み合わせることで物体の検出を行う。Deformable Part Model は現在までに、人や車、動物といったさまざまな物体に対して検出実験が行なわれているが、いずれも自然物体を対象としたものであり、コミック画像のような線画で表現された物体を対象とした実験は今までに行なわれていない。よって本論文では、コミック画像を対象として、Deformable Part Model を用いた検出を行い、その精度の評価を行った。

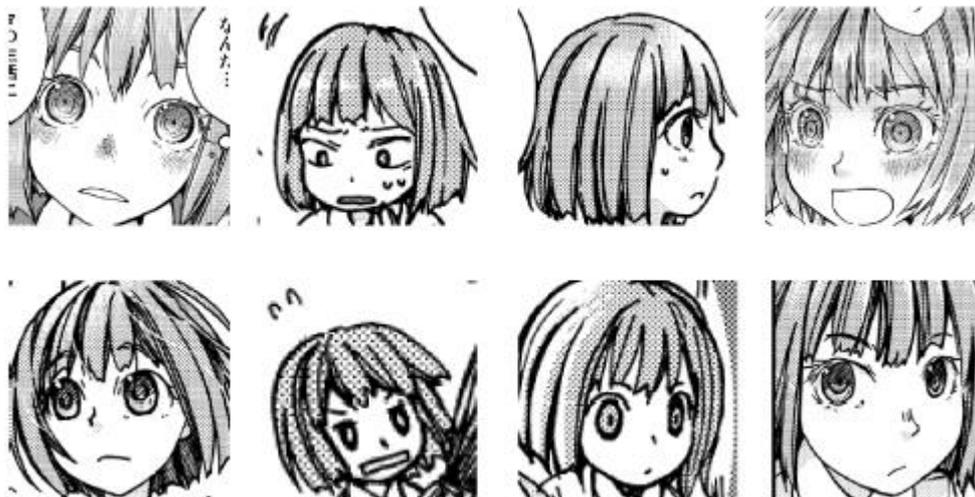


図 6 登場シーンにおける顔特徴の変化例(画像は[16]より)

3.2 実験の概要

Deformable Part Model を用いて、コミック画像内のキャラクター検出に関して実験を行う。具体的には、一つの漫画作品より登場人物の顔領域を指定し、学習を行なう。その後完成した識別器を用いて、学習サンプルに使用した既知の画像および学習サンプルに使用していない未知画像に対して、それぞれ人物の顔領域において検出が正しく行なわれているか検証を行った。本実験では Deformable Part Model の学習および検出に[17]を使用した。

3.3 実験条件

本実験で使用するコミック画像は、1 ページ 1342×1877 ピクセルのサイズのもの、26 ページ分とした。この中からランダムに抜き出した 9 ページを学習用サンプルとして使用した。本実験で検出対象とする顔領域は以下の三つの条件を満たしているものとした。

1. 目から顎までの範囲が映っている。
2. 両目が描かれている。
3. 60×60 ピクセル以上のサイズである。

検出対象の例を図 7 に、検出対象として含まない例を図 8 に示す。



図 7 検出対象の例(画像は[16]より)

本実験では、すべて漫画家の木野陽様 <http://www.etheric.com/> より学術目的のために許可を頂いたコミック画像を使用している^[16]。

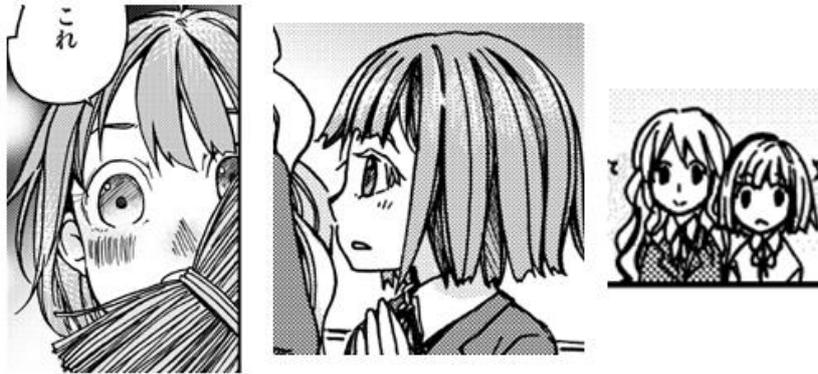


図8 検出対象として含まない例(画像は[16]より)

ポジティブサンプルには、コミック画像1ページ分の中から対象となる顔領域をバウンディングボックスで指定したものを使用した。ポジティブサンプルの例を図9に示す。ネガティブサンプルには、1ページに含まれている全てのコマの中から顔部分を含まない領域をそれぞれ切り出した画像を使用した。ネガティブサンプルの例を図10に示す。この結果、ポジティブサンプルは画像9枚のうち29箇所、ネガティブサンプルは画像89枚となった。

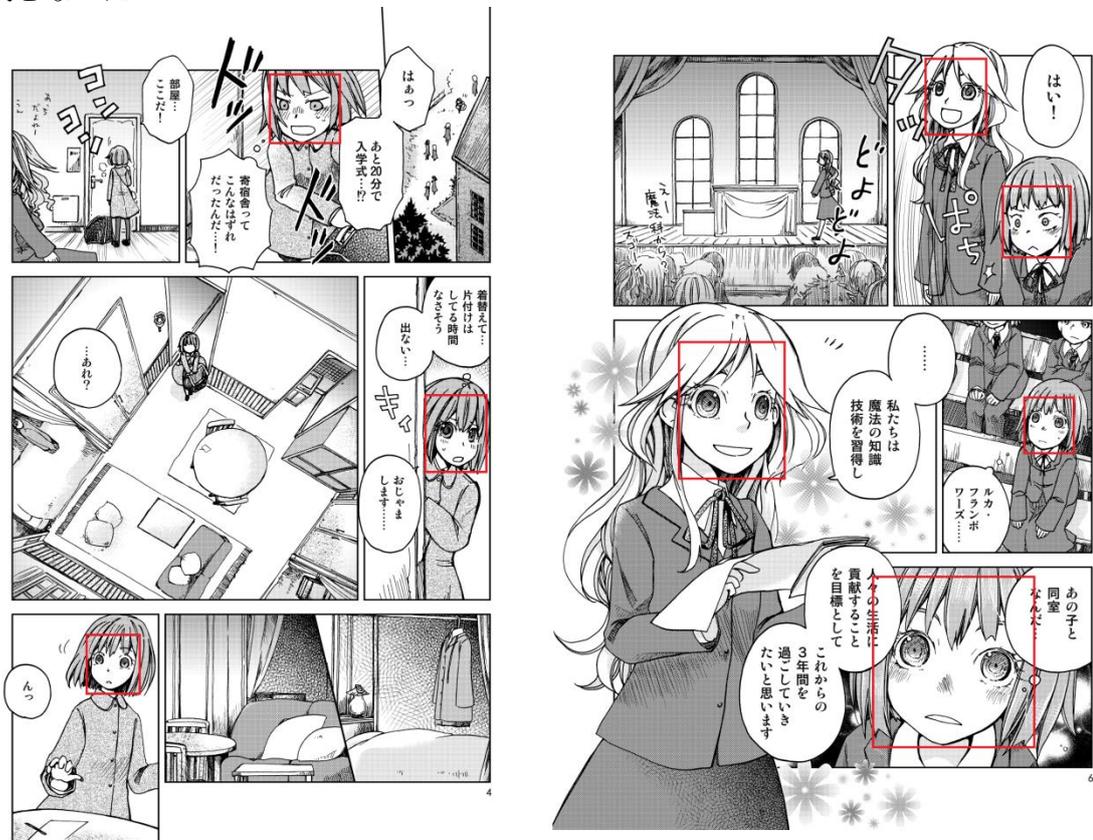


図9 ポジティブサンプルの例 (赤枠はバウンディングボックスで指定した領域。画像は[16]より)

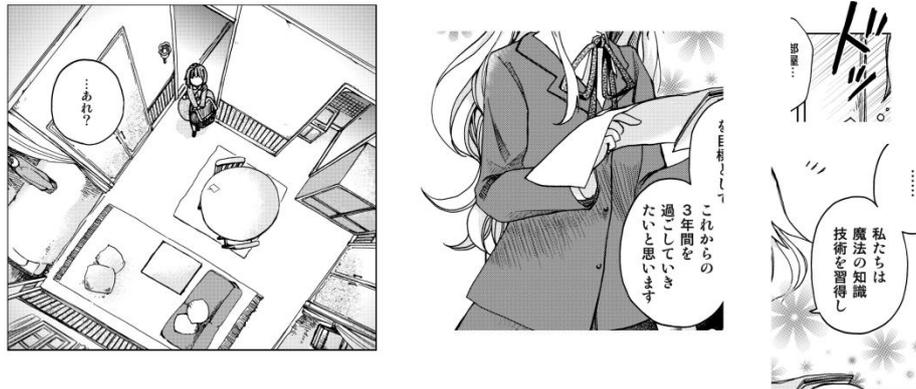


図 10 ネガティブサンプルの例(画像は[16]より)

3.4 実験結果

既知画像および未知画像について検出を行なった結果の例を図 11 に示す. また, Deformable Part Model によって生成された顔領域の検出モデルを図 12 に示す. 実験結果画像より, 登場人物の顔領域に対して検出が行なわれていることが分かる.

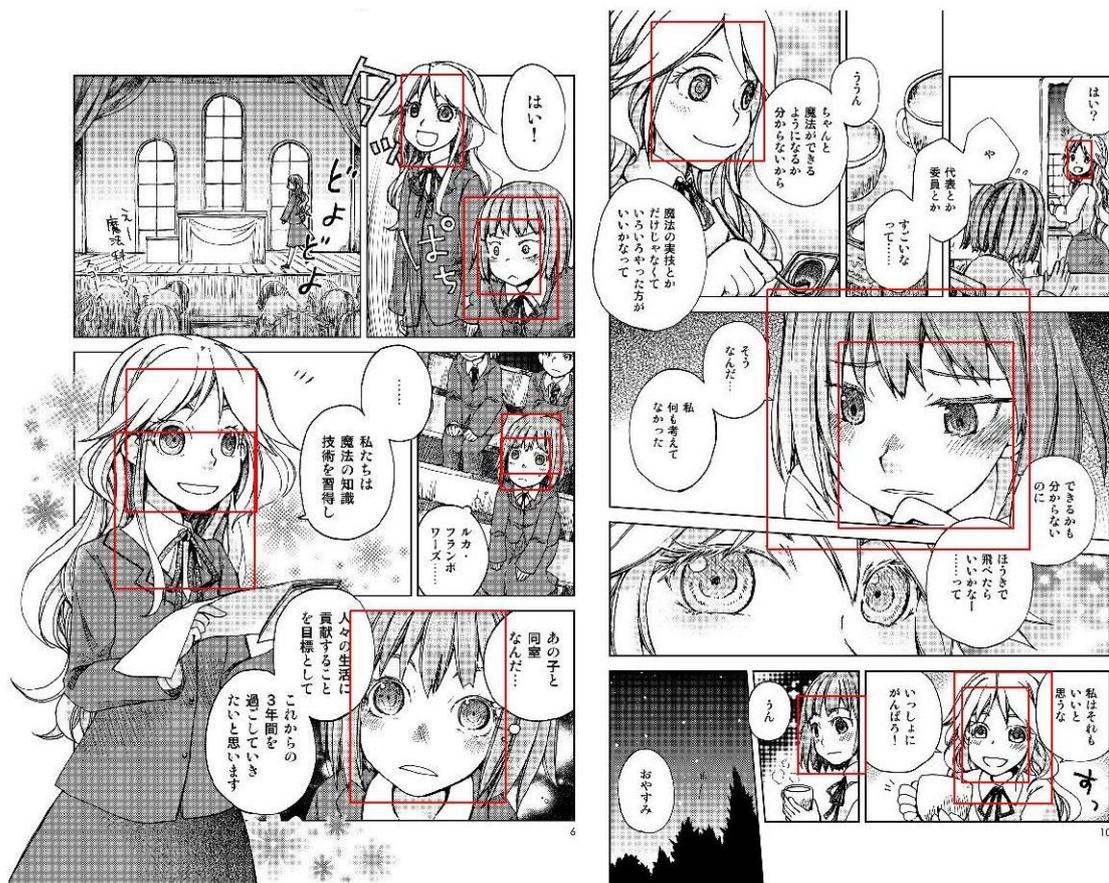


図 11 検出結果の例 (赤枠が顔部分として検出された箇所. 左側が学習に使用した既知の画像, 右側が未知の画像. 画像は[16]より)

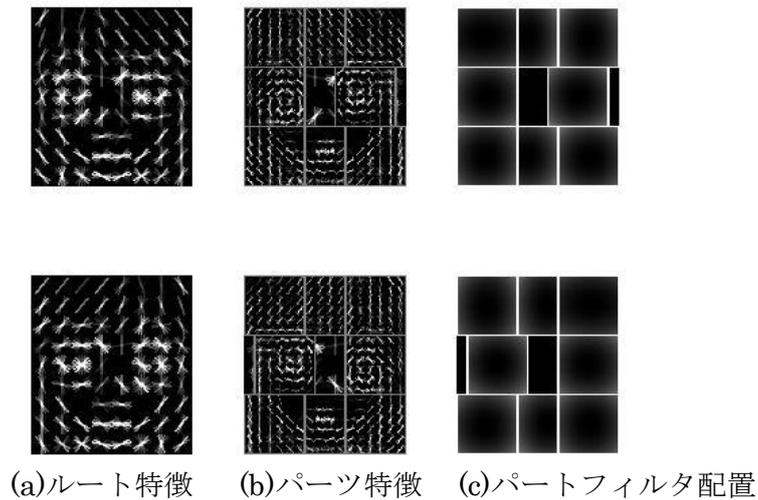


図 12 顔領域検出モデル

未知画像 9 ページ，既知画像 15 ページに対して検出処理を行なった結果を表 1,2 に示す．対象となる顔画像は，既知画像で 28 個，未知画像で 62 個であり，そのうち検出されたものは，既知画像 25 個，未知画像 57 個であった．この結果より，顔画像の検出率は既知画像で約 89.3%，未知画像では約 91.9%となった．

顔領域として検出された領域の数は，既知画像で 47 個，未知画像で 76 個であった．このうち顔領域を全て含んでいるものを全検出，顔領域を一部含んでいるものを一部検出，顔領域を全く含んでいないものをその他とする．図 13 に示す例では，全検出が 3 個，一部検出が 1 個となる．全検出である領域は，既知画像で 36 個，未知画像で 74 個であった．一部検出である領域は，既知画像で 11 個，未知画像で 2 個，その他の領域は既知画像，未知画像ともに 0 個であった．この結果より，検出された領域の正解率は，既知画像で 76.6%，未知画像で 97.3%となった．

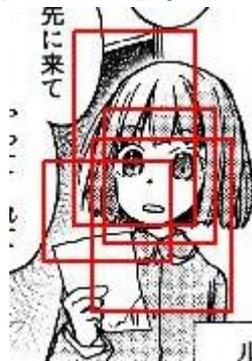


図 13 顔検出例(画像は[16]より)

表 1 検出された顔画像の個数

	対象となる顔画像	検出された顔画像
既知画像	28	25
未知画像	62	57

表 2 正しく検出された領域の個数

	検出した領域	完全検出	一部検出	その他
既知画像	47	36	11	0
未知画像	76	74	2	0

第4章 考察

第3章における実験の結果では、未知画像に対する検出率、正解率がともに90%を超えていた。また、検出した領域は、既知画像、未知画像ともに、全てが顔画像の一部を含んでいた。このことより、Deformable Part Modelがコミック画像においても物体検出が可能であることが示された。

既存のHOG特徴量とSVMを用いた手法と比較すると、既存手法では顔や瞳といった対象となる領域以外に、背景や吹き出しといった対象とは異なる領域について誤検出が存在したことに對して、今回の手法では、検出された領域は全て、顔領域の近傍に存在していることから、顔領域とそれ以外の領域の判別について、より正確な判別を行なっているといえる。よって、コミック画像における物体検出において、パーツ配置の特徴を利用することが検出精度を向上に繋がるということが示された。

一方、従来の検出実験では、未知画像よりも既知画像に對しての結果の方が良い値を示す傾向にあるが、今回得られた実験結果では、既知画像よりも、未知画像に對しての結果の方が検出率、正解率ともに高い値となった。このような結果になった原因として、以下の二つの問題が関連している。

1. 1つの顔部分に對して複数の検出が行なわれるケースが多い。

2. 一部の顔部分において、検出対象に該當するにも関わらず、検出されなかった領域が存在する。

一つ目の問題に關して、図14に例を示す。本実験結果では、一つの顔部分に對して顔部分の近傍が複数検出されるケースが多く見られた。また、未知画像よりも、既知画像の顔部分に對しての方が、近傍の領域を多く検出する傾向にあり、これが既知画像の正解率の減少に繋がっている。この問題の原因としては、一つに、Deformable Part Modelの特性から、検出される物体の形状変化に寛容なため、顔部分の近傍に對しても変化した顔部分と見なされ、検出してしまっていることが考えられる。また、今回ポジティブサンプルを与えるにあたって、顔領域が顔のみを含むか頭全体を含むかの定義が曖昧であったため、正解として検出する領域が不安定だったのではないかと考えられる。この問題の解決法としては、学習用画像の枚数を増やし、顔領域と近傍領域の違いを学習させることと、顔領域の定義をより明確にすることが挙げられる。

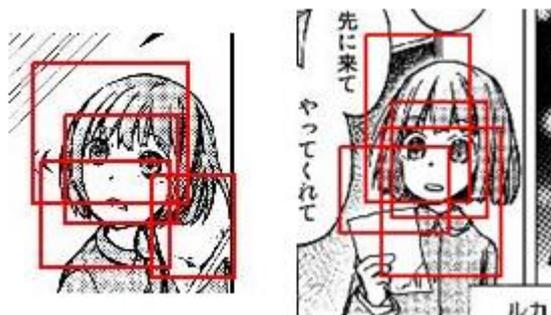


図14 顔領域の近傍が検出される例(画像は[16]より)

二つ目の問題に關して、図15に例を示す。今回、検出対象に該當するにも関わらず一部の顔部分に對しては検出が既知画像、未知画像ともに検出が行なわれなかった。これらの画像に對して、既知画像の方が、検出対象全体に對する割合が多かったため、検出率の減少に繋がった。これらの顔画像に共通しているのは、顔の特徴が他の顔画像と比べて違いが大きいことである。ポジティブサンプルとして与える数を増やすことで、これらの顔画像に對しても検出が行なえるようになると考えられる。よって、コ

ミック画像から安定した検出を行なうためには、キャラクターや表情ごとに一定数以上のポジティブサンプルを用意する必要があるといえる。



図 15 未検出の顔画像の例(画像は[16]より)

第五章 結論

本論文では, Deformable Part Model がコミック画像について 80%以上の顔画像に対して検出が可能であり, 最大で 97.3%の正解率で検出が行なえることを示した.

この実験結果より, Deformable Part Model はコミック画像に適用した場合についても物体検出が可能であることが分かった.

また HOG 特徴量のみを用いた手法との比較より, パーツの配置特徴を利用することは, コミック画像からの顔検出について効果的であることが分かった.

今後の課題として, コミック画像の学習サンプルに使用する画像について検討が必要である.

謝辞

本研究の機会及び素晴らしい実験環境を与えて下さり、貴重な時間を割いてご指導頂いた渡辺裕教授に心から感謝いたします。

本研究において、貴重な御意見や、御指摘を頂いた石井大祐助手に心から感謝いたします。

日頃から研究のアドバイスや議論をして頂いた研究室の皆様に心から感謝いたします。

本研究を行なうにあたって、コミック画像の提供および論文への掲載を許可いただいた木野陽様に心から感謝いたします。

最後に、私をここまで育てて下さった家族に深く感謝します。

平成 26 年 2 月 7 日

参考文献

- [1] インプレスビジネスメディア, “電子書籍ビジネス調査報告書 2013”, 株式会社インプレスビジネスメディア, 2013.
- [2] インターネットメディア総合研究所(編), “電子コミックビジネス調査報告書 2012”, 株式会社インプレスビジネスメディア, 2012.
- [3] 松下光範, “コミック工学の可能性”, pp.63-68, 2013.
- [4] 石井大祐, 河村圭, 渡辺祐, “分割線選択によるコミックのコマ分割に関する検討”, 情報科学技術フォーラム一般講演論文集, Vol.5, NO.3, pp.263-264, 2006.
- [5] 石井大祐, 河村圭, 渡辺祐, “コミックのコマ分割処理に関する一検討”, 情報処理通信学会研究報告, Vol.2012-AVM-76, No.1, pp.1-5, 2012.
- [6] Tanaka, T., Shoji, K., Toyama, F. and Miyamichi, J. “Layout Analysis of Tree-Structured Scene Frames in Comic Images,” Proc. 20th International Joint Conference on Artificial Intelligence , pp.2885-2890, 2007.
- [7] 野中俊一郎, 野沢拓也, 羽場典久, “コミックスキャン画像からの自動コマ検出を可能とする画像処理技術「GT-Scan」の開発”, FUJIFILM RESERCH & DEEVELOPMENT, No.57, pp.46-49, 2012.
- [8] 田中孝昌, 外山史, 宮道壽一, 東海林健二, “マンガ画像の吹き出し検出と分類”, 映像情報メディア学会誌, VOL.64, No.12, pp.1933-1939, 2010.
- [9] Dalal, N, Triggs, B, “Histograms of Oriented Gradients for Human Detection,” IEEE CVPR, pp.886-893, 2005.
- [10] 新井俊宏, 松井佑介, 相澤清晴, “漫画画像からの顔検出”, 電子情報通信学会総合大会, p.161, 2012.
- [11] 石井大祐, 渡辺祐, “マンガからの自動キャラクター位置検出に関する一検討”, 情報処理学会研究報告, Vol.2012-AVM-76, No.1, pp.1-5, 2012.
- [12] P.E Felzenszalb, et al., “Object Detection with Discriminatively Trained Part Based Models”, PAMI, vol.32, No.9, pp.1627-1645, 2009.
- [13] P.E Felzenszalb, et al., “A Discriminatively Trained, Multiscale, Deformable Part Model,” CVPR, 2008.
- [14] 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”, 電子情報通信処理学会信学技報, Vol.112, No.198, pp.1-14, 2012.
- [15] 三井相和, “CV 輪講 A Discriminatively Trained, Multiscale, Deformable Part Model”, FLAB Work Document Jul 15, 2008, pp.1-10, 2008.
- [16] 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.
- [17] Girshick, R. B. and Felzenszwalb, P. F. and McAllester, D., “Discriminatively Trained Deformable Part Models, Release 5”,
<<http://people.cs.uchicago.edu/~rbg/latent-release5/>>, <最終アクセス 2014/2/7>.

図一覧

図 1	HOG 特徴量を用いた学習による検出結果	4
図 2	HOG 特徴量	5
図 3	Deformable Part Model の人検出モデル	6
図 4	イメージピラミッド	6
図 5	Deformable Part Model のスコア算出の流れ	8
図 6	登場シーンにおける顔特徴の変化例	12
図 7	検出対象の例	12
図 8	検出対象として含まない例	13
図 9	ポジティブサンプルの例	13
図 10	ネガティブサンプルの例	14
図 11	検出結果の例	14
図 12	顔領域検出モデル	15
図 13	顔検出例	16
図 14	顔領域の近傍が検出される例	17
図 15	未検出の顔画像の例	18

表一覧

表 1	検出された顔画像の個数	16
表 2	正しく検出された領域の個数	16