情報処理学会研究報告
IPSJ SIG Technical Report

マンガからの自動キャラクター位置検出に関する検討

石井大祐† 渡辺裕†

近年画像特徴量を用いた画像認識技術の発展は顕著である．これまで画像認識では，主に自然画像を対象とした特徴抽出，記述および解析が行われてきた．一方，電子書籍および電子コンテンツの領域では，電子化されたマンガが一般的になりつつある．電子コンテンツの利便性を高める上で，マンガにおける内容理解技術は重要である．本稿ではマンガを対象としたキャラクターの検出手法について検討し，HOG 特徴量および SVM を利用したキャラクターが存在箇所の検出を試みた．結果としてキャラクターの瞳部分を学習に使用した際に顔全体の画像を学習に使用した場合と比較して，誤検出を抑えつつキャラクターの顔位置が検出されることを確認した．

A Study of Automatic Human Detection for Comic Image

Daisuke Ishii† and Hiroshi Watanabe†

Recent years, glowing of image analysis technology is remarkable. Feature detection, description and analysis methods for natural image are mainly researched in image analysis region. On the other hand, the electronic comics are growing popular widely in the electronic publishing and electronic content area. To improve the usability of electronic contents, it is important that the analysis technologies for comics. In this paper, we study the human detection method for comic images and attempted to detect human existing areas by method that uses HOG feature and SVM. As a result, we have confirmed that the method trained by the eye training set can obtain better result than trained by the face training set.

1. はじめに

マンガはクールジャパンとして国内でも重要なコンテンツとして認識されている．知的財産戦略本部コンテンツ・日本ブランド専門調査会による日本ブランド戦略[1]においても，マンガについて，文化資源としてのアーカイブ，日本ブランドとしての海外展開，およびこれに伴う模倣品や海賊版対策等について触れられている．また，近年の電子書籍市場の急速な発展においても，マンガはその重要な位置を担うコンテンツである．電子書籍ビジネス調査報告書[2]によると，マンガのユーザー数およびコンテンツ数は急速に拡大しており，市場全体に占める割合も大きい．このように，マンガは国策および電子書籍市場において大変重要な位置を占めるコンテンツである．

ディジタル化されたマンガ画像のアーカイブにおいて，マンガの内容に関するメタデータを抽出し，さらに重畳し元のデータと併せて保存をしておく事で，より利便性の高いアーカイブを作成する事が可能となる．たとえば，アーカイブからの検索において，マンガのタイトルや作者情報だけでなく，マンガ内のキャラクター，台詞，シーンなどの情報を記録し，利用することで，特定のキャラクターの登場するシーンの検索や提示など，よりユーザセントリックな検索機能を提供する事が可能となる．

マンガのキャラクター認識に相当する技術は，現時点ではまだ確立されていない．もっとも近い分野として自然画像を対象とした画像解析の分野があり，人物，顔の検出に関連する研究が行われてきた．一般の顔画像検出では Scale Invariant Feature Transform (SIFT)[3]や Histograms of Oriented Gradient (HOG)[4]などの特徴量を利用した解析処理が実現されている．これらの技術は，ディジタルカメラ等の顔検出処理などを始めとする様々なアプリケーションに利用されている．これらの手法で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高性能な判別器を構成している．

マンガは自然画像と異なる信号的特徴を持ち，かつ自然界に存在しない人物等が描かれる．このため，自然画像用のアルゴリズムをそのまま使用しても，良好な解析結果は得られない．マンガは主に以下の 3 つの組み合わせにより描かれている：1.白黒の 2 値による線画． 2.塗りつぶ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．マンガ画像は自然画像と比較して，画像中の信号としてエッジ成分と平坦領域が多く含まれる．マンガを対象とした解析処理を実現する為には，マンガ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．

本稿では，マンガからの登場人物が存在する位置の自動検出を目的とした，HOG 特徴量および SVM による識別機の構成とスライドウィンドウを利用した画像からのオブジェクト検出処理について検討を行った．

† 早稲田大学大学院国際情報通信研究科
Graduate School of Information and Telecommunication Studies WASEDA University

本稿の構成は以下のとおりである．2 章ではマンガ画像解析関連技術として，マンガの画像的特徴，マンガ解析の一例であるコマ分割処理の概要および今回利用する HOG 特徴量について述べる．3 章では，HOG 特徴量と SVM およびスライドウィンドウによる画像内探索を用いたマンガ画像内のオブジェクト検出処理に関して述べる．4 章では，3 章にて述べたオブジェクト検出処理を利用したキャラクター検出実験について述べる．5 章で本稿をまとめる．

2. マンガ画像解析関連技術

2.1 一般的な画像解析

近年画像特徴量を利用した画像解析技術が多数提案されている．とりわけ，ディジタルカメラ等では，リアルタイムの人物顔検出処理が行われ，実用的な面でも利用価値の高い検出処理が実現されている．このような処理では，SIFT[3]特徴量などを利用し，多数の人物の顔を学習させることで，自然画像中の人物の顔部を高速に特定可能とである．また，監視処理などを目的として，HOG[4]特徴量を利用し，人物の輪郭特徴をとらえ，これにより高い精度の画像内人物検出処理を行う方法が提案されている．

これらの手法では，自然画像中の人物の特徴を，その利用形態に応じた検出対象である顔や人体全体などに合わせて，最適な特徴量を利用し解析を行う．

2.2 マンガの特徴

マンガは，主に白黒の２値による，線画，塗りつぶし領域，白黒の点密度を変化させてグラデーション表現を行うトーン等のテクスチャ領域，台詞などの文字領域によって構成されている．マンガのページには構成要素（登場人物，視覚効果，テキスト等）がコマと呼ばれる（多くの場合四角形に区切られた）領域上に配置され，読者はこのコマの並びに沿ってマンガを読む．輝度の変動としてマンガ画像をとらえると，マンガではほぼすべての情報がエッジによって描かれているとみなすことができる．これは自然画像とは異なるマンガ画像特有の信号的特徴であり，一般的な画像解析処理をマンガ画像で行うことが困難な原因の一つと考えられる．

2.3 コマ分割処理

我々はこれまでにマンガのコマの位置および並び順の情報を得るために，コマ分割処理について提案を行ってきた[5]．コマ分割処理では，画像の２辺を結ぶ直線上に高強度のエッジ成分が重なる場合に，この直線を基準とした画像の２分割処理を行う．この 2 分割を繰り返し行うことでマンガ 1 ページを各コマに分割する．我々は，この手法において，直線上のエッジ方向の分布を調べることで，コマ分割成功率を改善した．この検討から，マンガ画像においてはエッジの方向性が有効な識別規範となる可能性が示唆される．

2.4 HOG 特徴量

HOG 特徴量は，画像の特徴量記述子として，特にエッジの方向についての特徴を利用した特徴量である．HOG 特徴量は一般的に以下の手順により求められる．

１．画像全体を任意サイズのセルに分割する．
２．セル内の各画素において，近隣画素間の輝度こう配量およびその方向を求め，方向ごとにこう配強度を加算する．ここで方向数は任意の値とする．
３．連続した複数のセル（例えば縦 3×横 3 の合計 9 セル）を集めてブロックを構成し，ブロック内で加算されたこう配強度の正規化を行う．
４．ブロック内のセルごとに，方向の数に対応する正規化されたベクトルを HOG 特徴量の一つのベクトルとする．
５．ブロックを 1 セル分ずつスライドさせていき，画像中のすべての領域において３および４の処理を行う．

HOG 特徴量の次元数は 1 ブロックに含まれるセルの数×画像内のブロック数×方向数 = 次元数として算出される．例として入力画像が 120pixel×120pixel セルのサイズを 20pixel×20pixel，1 ブロックを 3×3 セル，方向数を 5 とすると，この場合の HOG 特徴量の次元数は 9×16×5 = 720 次元となる．

3. マンガ画像内オブジェクト検出処理

3.1 マンガ画像内オブジェクト検出処理の概要

今回オブジェクト検出に利用した処理は以下に示す 3 つの処理からなる： 1. 学習用正例，負例の用意 2.特徴量の取得 3.SVM による学習および判別.

初めに検出対象の正例，負例画像を切り出しておく．切り出された画像群に対して画像を一律解像度に変換し特徴量の算出を行う．得られた特徴量群に対してサポートベクタマシン(SVM)による学習を行う．判別過程ではスライドウィンドウを用いて多数の小領域を取り出し，小領域ごとに特徴量抽出と判別処理を行う．

3.2 特徴量

マンガ画像は 2.2 節で述べたように，多数のエッジの集合体により表現される画像である．また，エッジの方向性を利用することで，マンガのコマ分割を最大で 90% 以上の精度で実現できることから，エッジの方向性を利用することは，マンガの解析を行う上で有効であると考えられる．エッジの方向性を利用した判別を行うためには，エッジ強度および方向性の分布を特徴量とする HOG が適当である．そこで本検討では，画像特徴量として HOG 特徴量を採用した．SVM にて規範となる HOG 特徴量の学習後，得られた識別器を用いて入力画像の判別処理を行う．

SVM の学習には Lin らによるライブラリ libsvm[6]を利用した．SVM には C-SVM を利用し，その他のパラメータはデフォルト値を用いた．

3.3 スライドウィンドウによる検出対象画像内走査と判別処理

入力画像中の局所領域における HOG 特徴量を調べるため, スライドウィンドウによる切り出しを行い, 各画像を一律の解像度に変換した後で, 特徴量算出と SVM の学習結果による判別を行う．マンガには 1 画像中に多数の要素が配置されているため，入力画像を局所領域に分割して各々特徴量を求める必要がある．オブジェクト検出処理では，対象画像からスケールを変化させつつ，スライドウィンドウにて多数の小領域画像群を取得し，それぞれに対して HOG 特徴量の算出と，SVM による判別を行う．

スライドウィンドウは, 指定された任意サイズ(S)の正方形からその N 倍となるまで，ウィンドウの一辺の大きさ（スケール）を 2 倍ずつ拡大しつつ，画像内からウィンドウサイズの画像の切り出しを行う．各ウィンドウの移動量は処理対象のウィンドウサイズの K 倍とする． 上記ルールに従って，調査対象画像 1 枚に対して得られる小領域画像数は, 対象画像サイズを 1111×1554[pixel], S=20[pixel], N=4, K=0.5 とした場合に 21740 となる．

上記の走査により得られた各小領域画像を，学習用画像群と同じ画像解像度に変換する. 次に, 2.4 節に記載した方法に基づき各小領域画像毎に HOG 特徴量を算出する．算出された HOG 特徴量を学習済み SVM により判別する．判別処理では 1 もしくは-1 の 2 値の判別結果を出力する．判別処理により 1 と判定されたウィンドウについて，検出対象がそのウィンドウに含まれていると判断する．

4. マンガ画像内キャラクター検出実験

4.1 実験の概要

3 章にて述べたマンガ画像内オブジェクト検出処理を利用して，マンガ内のキャラクター検出に関して実験を行った．具体的には登場人物の顔の一部である瞳または顔部分全体を学習し，完成した判別器を用いて，人物の顔部分において判別が正となる検出結果が得られるかを測定した．

4.2 実験条件

本検討では，登場人物（人物の顔）位置を検出することを目的としているが，顔の瞳部分に関しては多くのマンガにおいて特徴的な書き込み（ベタ塗りやトーン等を利用した, 他の部分と異なる瞳特有の表現）が見られるため, 有意な特徴を持つと考え，これを学習用の正例画像として採用した．また，比較のためキャラクターの顔全体も学習用の正例画像として採用した．負例には顔，瞳以外の部分をランダムに抽出した画像を利用した．学習に利用した画像を図 1 から図 3 にそれぞれ示す．

図 1　学習用画像群 [正例 1 : 瞳] （一部抜粋）瞳部分の学習に用いる正例として，瞳部分を含む用にランダムなサイズで切り出しを行った画像合計 53 枚を使用した．

図 2　学習用画像群 [正例 2 : 顔] （一部抜粋）顔部分の学習に用いる正例として，顔部分全体を含む用にランダムなサイズで切り出しを行った画像合計 70 枚を使用した．

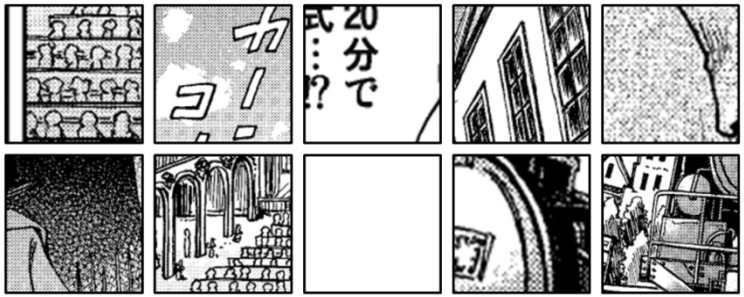

図 3　学習用画像群 [負例 : 顔以外] （一部抜粋）負例の学習画像は顔および瞳を含まない箇所についてランダムサイズに切り出しを行った画像合計 113 枚を使用した．

図 4　検出結果：瞳を正例として学習した識別器による画像内探索処理結果．SVM により正と判断されたウィンドウについて赤色の四角形で該当箇所を示した．

図 5　検出結果：顔全体を正例として学習した識別器による画像内探索処理結果．

瞳の正例には 53 枚，顔の正例には 70 枚，負例には 113 枚の画像をそれぞれ使用した．また，一枚の画像から取得可能な正例数に限りがあるため，切り出し位置およびサイズを変化させるなどして一つの瞳などから複数の正例画像の切り出しを行った．ただし，ローテーション処理による正例の増加は行っていない．すべての画像は一意の画像サイズ 50×50[pixel]に変換して特徴量の抽出を行った．HOG 算出時のパラメータは各セル 10×10[pixel]，各ブロックは 3×3 セル，方向成分の数は 9 とした．HOG 特徴量の次元数は 729 次元である．検出ウィンドウの最小サイズは 20×20[pixel]とし，160pixel までの 4 段階についてスライドウィンドウによる走査を実施した．検出対象入力画像サイズは 1111×1554[pixel]で，各画像における総ウィンドウ数は 21740 である．

4.3 実験結果

瞳もしくは顔を正例とした場合の実験結果の例を図 4 および図 5 にそれぞれ示す．結果画像において，左側は学習に使用した箇所が含まれている画像，右側は学習では使用していない未知画像に対する処理結果である．画像中の赤い四角領域は検出対象が含まれていると判別されたウィンドウの位置およびスケールを示す．

実験結果画像より，瞳と顔のどちらの場合においても，登場人物の顔部分に検出対象があると判別されていることがわかる．一方，文字部分が多く検出されていることが確認された．ただし，瞳を正例として学習を行った場合の方が，顔を正例とした場合と比較して文字を初めとする顔以外の部分における検出数が少ない．従って瞳部分がマンガ画像内において他の部位と異なる特徴的なエッジの特徴を持つことが推察される．

なお，本検討では，マンガの解析に有効な特徴量についての調査段階である為，顔領域の正確な定義まで言及できておらず，現時点では客観評価を行っていない．

図 6　検出結果：図 4 と同様の処理（正例を瞳とした場合）において，検出ウィンドウサイズを 40 - 160px に制限した場合の検出結果.

4.4 ウィンドウサイズによる検出結果への影響

今回の入力画像では，文字サイズと同程度に小さい登場人物はほぼ存在しないため，検出ウィンドウの最小サイズを 40pixel に変更して実験を行った．処理結果を図 6 に示す．ここでは瞳を学習の正例としている．図 4 と比較して，瞳（人物）部分の検出結果を残しつつ，文字部分の誤検出が減少している．画像の特徴について，文字サイズと顔のサイズの比率など，既知の情報を加味することで，検出結果の改善が得られる場合がある．このことから，ターゲットとなる入力画像の特性が事前に推測できる場合には，適切にウィンドウサイズを制限することで検出精度を向上させることが可能である．

5. おわりに

本稿では，マンガ画像上においてキャラクターが存在する箇所を抽出するために，HOG および SVM を用いた判別器を生成し，これを利用したキャラクター位置の検出を実験した．

実験では，キャラクターの瞳部分および顔部分を正例，それ以外の箇所からランダムに選択した場所を負例として判別器を生成し，それぞれ該当領域の検出処理を実施した．その結果，どちらの判別器においても，キャラクターの顔部分において正の検出結果を得ることができた．とりわけ，瞳を正例とした判別器では，顔を正例とした物と比較してより偽である領域の検出数が少ないことが確認された．

また，画像の既知の情報を利用し，検出処理時のウィンドウサイズに制限を設けることで，文字領域に発生する検出数を削減する効果があることが確認された．

今回の検討では HOG 特徴量を単体で利用した検出処理を行ったが，今後の課題として，複数の特徴量（テクスチャのパターンなど）を併せて学習することで，より高精度な解析処理を実現可能と考えられる．

注) 図 1 から図 6 の画像については文献[7]の画像をもとに著者らが加工を行った物である．

謝辞　本稿では，マンガ家の木野陽様 http://www.etheric-f.com/ より学術目的の為に使用許可を頂いたマンガを実験に使用した．マンガ画像の提供および原稿への掲載を許可いただいた木野陽様に深く感謝する．

参考文献

1) “ソフトパワー産業を成長の原動力に～（平成 21 年 3 月 10 日 内閣官房 知的財産戦略本部）” http://www.kantei.go.jp/jp/singi/titeki2/houkoku/090310_nihonbland.pdf 最終確認 2011 年 10 月 18 日
2) 高木ら，”電子書籍ビジネス調査報告書，”三橋昭和，インプレス R&D，東京，2009.
3) Lowe, David G. (1999). ″Object recognition from local scale-invariant features,″ Proceedings of the International Conference on Computer Vision. 2. pp. 1150-1157.
4) Dalal. N, Triggs. B, “Histograms of Oriented Gradients for Human Detection,” IEEE CVPR, pp. 886-893, 2005.
5) 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.
6) 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.
7) 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.