

マンガ画像における瞳検出の解像度依存性に関する検討 A Study for Resolution Dependency of Eye Detection on Comic Image

石井 大祐[†] 渡辺 裕[†]

Daisuke ISHII[†] and Hiroshi WATANABE[†]

[†] 早稲田大学国際情報通信研究科

[†] Global Information and Telecommunication Studies, WASEDA University

E-mail: [†] d-ishii@fuji.waseda.jp

概要

マンガにおける重要な要素の一つとしてキャラクターがある。マンガ画像からキャラクターの位置を検出する為に、瞳の検出が有効である。瞳領域は、顔領域と比較してオクルージョンが少ないためである。また、瞳部分はマンガの他の部分と比較して特徴的な書き込みによって描かれている。一方で、瞳部分の書き込みは細かく、画像の入力解像度に依存してその特性が変化する。本研究では、高精度な検出処理の実現を目指し、入力解像度が、瞳部分の検出処理に与える影響について検討する。

1. はじめに

近年、コンテンツとしてマンガへの注目が高まっている。とりわけ、マンガはアニメーションと並び、国外からも高い評価を得ている。また、国内でも、電子書籍市場において、その大部分をマンガが担っている[1]。このように、マンガは多数のユーザーにとって魅力的なコンテンツであり、近年隆盛を極める電子書籍においても大変重要なものである。

画像をデジタル化し、アーカイブする際に、様々なメタデータが付与される。たとえば、写真ではその撮影条件として、カメラパラメータに関する情報や、GPSシステムを利用した撮像位置情報を加えて保存しておくことで、撮影後の画像に対する利便性を高めている。また、近年では被写体となっている人物の自動認識等も行われている。動画画像でも、内容解析による、要約映像生成など、コンテンツの中身に関する情報を抽出し、これをメタデータとして付与することで、コンテンツの利便性が高められている。

マンガは多数の要素によって表現されている。ページ上に読む領域や順序の目安となる、コマと呼ばれる枠があり、その中に登場人物・セリフ・背景・視覚効果などを描くことで、読み手に対して、ストーリーと、それに付随する様々な場面を見せている。従って、マンガに含まれる情報としては、写真や動画画像と同様であり、登場人物や、内容の要約等の情報を付加する

ことで、マンガコンテンツの利便性を高めることができると考えられる。

マンガ画像は、自然画像とは異なる特徴を持つ。マンガに描かれる絵は、自然画像と異なり、作者の想像を基にして描かれる、自由度の高い画像である。さらに、基本的には白黒の線や点（スクリーントーン）、および塗りつぶしによって表現されており、デジタル化した際の信号の特徴も、自然画像とは異なるものである。従って、マンガ画像を解析する為には、マンガ画像特有の画像解析処理が必要である。

我々は、マンガ画像上から、登場人物の存在を検出するための手法について検討を行った。マンガにおいて、登場人物はほぼすべてのマンガに見られる要素である。また、写真画像や動画画像の解析においても、人物は重要な検出対象であり、メタデータとしても最も活用されている要素である。これまでに我々は、マンガから登場人物の存在有無を調べるために、登場人物の瞳を検出する手法を提案している[2]。瞳の検出では、画像特徴量として、Histogram of Oriented Gradient (HOG)、学習に Support Vector Machine (SVM)を利用している。マンガでは、画像の表現として、スクリーントーンや、大量の細い線などによって描かれる領域が多数存在する。これらは、解析を行う画像の解像度により、折り返しひずみの発生や、濃淡パターンの縮退などが発生する為、解析結果に影響を及ぼすことが考えられる。

本稿では、この検出画像および HOG 特徴量計算時に使用する画像の解像度が検出結果に与える影響について検討を行った。本稿の構成は以下の通りである。2章では、これまでに提案を行ってきたマンガ画像における瞳検出処理、およびその検出結果について述べる。3章では、2章で述べた瞳検出処理を用い、学習時および検査時に HOG 特徴量を計算する画像のサイズを変化させた場合の比較実験について述べる。4章で本稿をまとめる。

2. マンガ画像における瞳検出処理

2.1 瞳検出処理の概要

マンガから登場人物の存在を検知する為に、我々は、下記の理由から瞳部分に注目した検出処理を検討した。マンガでは、人物の顔全体が同一形状で描かれず、他の要素によって覆い隠される場合がある。また、登場人物の向きにより、自然画像では想定しえないような変化が生じる。これと比べ、瞳部分には、他の箇所にはない、特徴的な書き込みが多くみられ、また、その変異が顔全体を検出する場合と比較して小さい。従って、顔全体を検出しようとするよりも、瞳をターゲットとして検出処理を行った方が、その検出精度が高くなると考えられる。瞳が存在する場合には、その周囲に登場人物がいると判断することとした。

瞳検出処理では、下記に述べる HOG 特徴量を用い、以下の手順で判別を行う。初めに、学習用画像を任意の解像度に正規化し、HOG 特徴量を求め、SVM により学習し判別器を生成する。そして、検出対象画像を学習用画像と同様の解像度に正規化し、そこで得られる HOG 特徴量に対して SVM による判別器を用いて判別する。

2.2 HOG 特徴量

輝度こう配を利用した画像の特徴量として、HOG 特徴量がある[3]。我々は、マンガ画像に対する画像解析処理として、コマの並びおよび位置情報を取得する手法を提案してきた。その中で画像に含まれる輝度こう配方向の情報を利用することで、その検出精度が向上可能であった。従って、画像特徴量として、輝度こう配の方向を用いた HOG 特徴量は、マンガ画像の特徴を扱う上で有効であると考えられる。

一般的な HOG 特徴量は、次の手順により求められる。初めに、画像を任意サイズのセルに分割する。次に、セル内の各画素における隣接画素との輝度こう配量を求める。これを、輝度こう配方向を任意の数（9 程度）に離散化させ、その方向ごとに加算し、それぞれ合計を求める。このようにして求めたセルをまとめ（ 3×3 セル程度の）ブロックを構成する。ブロック内では加算された輝度こう配の総計をもちいて正規化を行う。このようにして求められた各セルにおける方向毎の値を HOG 特徴量の 1 ベクトルとする。ブロックを 1 セル分ずつスライドさせてゆき、画像中のすべての領域において上記ベクトルを求めることで、HOG 特徴量を構成するベクトル群が得られる。

マンガ画像から HOG 特徴量を求める際には、学習を行いたい箇所について、あらかじめ画像からの切り出しを行う。切り出された画像に対して、上記の手順にならない HOG 特徴量を求める。ここで、学習対象と

したい領域の大きさは、シーンによりそれぞれ異なるため、切り出される画像の画素数は不定である。

HOG 特徴量を学習、判別する為には特徴ベクトル数が一定である必要がある。ここでは、切り出された画像を任意の共通画素数($N \times N$)に解像度変換し、各セルの画素数、セル数、ブロック数がすべて同一となるように設定し、HOG 特徴量の算出を行う。

2.3 マンガ画像内瞳検出処理

マンガ画像から、瞳領域を検出する為に、スライドウィンドウを用いて、ウィンドウ内に含まれる画像に対してそれぞれ HOG 特徴量を計算し、あらかじめ求めた学習結果を用いて、瞳か否かの判別処理を行う。

スライドウィンドウによる走査では複数サイズのウィンドウを用いて走査を行う。これは、画像中のシーンにより、検出対象である瞳の大きさが異なるためである。各ウィンドウによって切り出される画像はすべて先の学習時の解像度($N \times N$)と同一となるように解像度変換を行い、HOG 特徴量を求める。例として、 2684×3754 画素（B5 サイズ 370dpi 相当）の画像に対してウィンドウサイズを 25×25 , 50×50 , 100×100 , 200×200 , 400×400 として捜査を行った場合、切り出される領域の総数は 81404 である。ただし、画像からはみ出す領域についてはウィンドウによる切り出しを行わない。

3. マンガ画像からの瞳検出実験

3.1 実験の概要

マンガは基本的に白黒の線及び、スクリーン等により描かれる。この画像的特徴により、解析を行う画像の解像度により、折り返しひずみの発生や、濃淡パターンの縮退などが発生する為、解析結果に影響を及ぼすことが考えられる。

検出画像および HOG 特徴量計算時の解像度が検出結果に与える影響について調査するため実験を行った。実験では、2.2 および 2.3 章で述べた、HOG 特徴量算出時に正規化を行う画像サイズ($N \times N$)を変化させることで、検出結果に対する影響を調べた。

3.2 実験条件

実験に使用した画像は 2684×3754 画素（B5 サイズ 370dpi 相当）である。これは、もともと 740dpi 相当でスキャンされた画像を 2 分の 1 サイズに再サンプリングして得た画像である。高解像度の画像を使用した場合には計算量が多分に大きくなることが予想されたため低解像度化を行った。しかし、これは通常電子書籍として配布されるマンガ画像と比較して十分に高解像度であり、スクリーン等には、低解像度化による折り返しひずみの影響がほぼ見られない解像度を確保した。

3.3 学習と判別処理

入力画像の半数から学習に用いる正例（瞳部分）



図 1 学習用正例画像（一部抜粋）

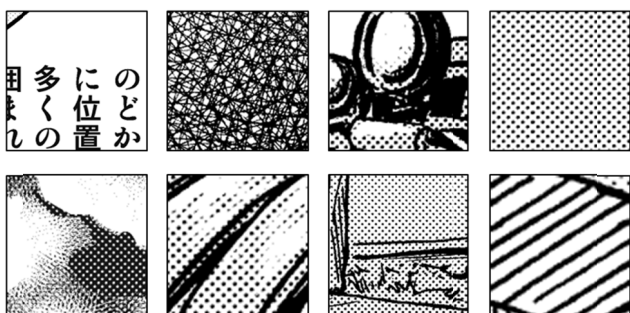


図 2 学習用負例画像（一部抜粋）

画像と、負例（瞳以外の部分）を切り出し、libsvm[4] による学習を行った。学習に使用する画像により、判別結果に影響が出ること確認されている。このため、今回は、一部の局所的な特徴に影響されにくい学習用画像を利用した。学習に用いた正例および負例画像の一部を図 1 および図 2 にそれぞれ示す。学習に用いた正例画像は 111 枚、負例画像は 185 枚である。

判別処理に用いる入力画像は、学習に用いた画像と同一解像度の画像を使用した。ウィンドウサイズは 25×25 , 50×50 , 100×100 , 200×200 , 400×400 の 4 種類とし、切り出される領域の総数は 81404 である。また、画像からはみ出す領域についてはウィンドウによる切り出しを行わない事とした。

3.4 実験と考察

学習時および、判別時に利用する正規化解像度 ($N \times N$) について、 $N=25, 50, 100, 200$ についてそれぞれ実験を行った。入力画像および各正規化解像度に対応する実験結果の画像を図 3 および図 4 に示す。判別対象の画像は図 3 が学習に使用した画像、図 4 が学習に使用していない未知の画像である。

$N=25$ の場合には、検出結果にトーン部分が多く含まれたと考えられる。これは正規化時に発生する折り返しひずみの発生と、テキストチャと書き込み部分の HOG 特徴量同一化が発生したためと考えられる。 $N=50$ 以上の場合には、トーン部分に対して検出が得られることはほぼ見られないため、上記ひずみ等の影

響を抑制できていると考えられる。一方で、 $N=200$ まで上げていくと、瞳内部の書き込み部分（複雑で多方向に分布している）と、他の箇所でも書き込みの多い部分が、HOG 特徴量を計算する過程で近い値となり、結果的に今回の実験では検出精度が低下する結果となったと考えられる。

4. まとめ

本稿では、HOG 特徴量を用いたマンガ画像における瞳検出処理の、解像度依存性に関して検討を行った。今回の実験では $N=50$ 程度が最も良好な結果を示すことが確認された。従って、本手法では、瞳の丸い部分と、瞳内部の書き込みから得られる適度な複雑さが緩和されている状態が、瞳の検出において、最も良好な結果が得られると考えられる。

今後の課題として、今回は HOG の次元数を固定とし、入力画像側の解像度変換を行ったが、一方で、HOG 特徴量の次元数のみを固定とし、解像度については一律に変換しない方法も考えられるため、検討が必要である。瞳の丸い部分は、線も太く描かれていることが多く、瞳を構成するうえで重要な情報であると考えられる。しかし HOG 特徴量などの輝度こう配のみを考慮する手法においては、この塗りつぶされた範囲の大きさは無視される。そこで、瞳の塗りつぶし領域を評価する方法を検討することで、より実用的な瞳検出処理を実現できる可能性が考えられる。また、今回利用した特徴量は、切り出した範囲によってその特徴量の傾向が異なるため、シフト不変性およびスケール不変性が高いとは言えない。このため、これらの特性を考慮した特徴量記述の方式を検討することが考えられる。

5. 謝辞

本稿にて提示したマンガは木野陽様 <http://www.ethheric-f.com/> より学術目的の為に使用許可をいただいた物である。（注）図 1 から図 4 の画像については上記画像を基に著者らが一部加工を行った。マンガ画像の提供及び原稿への掲載許可をいただいた木野陽様に深く感謝する。本研究は科研費 24700100 の助成を受けたものである。

文 献

- [1] 高木ら, 電子書籍ビジネス報告書, 三橋昭和, インプレス R&D, 東京, 2009.
- [2] 石井大祐, 渡辺裕 “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,” 情報処理学会研究報告, 2012-AVM76(1), pp.1-5, Feb. 2012.
- [3] Dalal. N, Triggs. B “Histograms of Oriented Gradients for Human Detection,” IEEE CVPR, pp. 886-893, 2005.
- [4] 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.



3-(A)



3-(B)



3-(C)



3-(D)



3-(E)

図 3 瞳検出実験結果 (学習に使用した領域が含まれる) A:入力画像,B:検出結果 N=25, C:検出結果 N=50, D:検出結果 N=100 E:検出結果 N=200 赤色の領域は SVM による判別結果が正となった箇所を示す。



4-(A)



4-(B)



4-(C)



4-(D)



4-(E)

図 4 瞳検出実験結果 (学習に使用した領域が含まれない) A:入力画像,B:検出結果 N=25, C:検出結果 N=50, D:検出結果 N=100 E:検出結果 N=200 赤色の領域は SVM による判別結果が正となった箇所を示す。