

# 画像認識とマイクロタスク型クラウドソーシングを組み合わせた マンガのコマ領域の判定

## Hybrid Identification of Frame Areas in Manga using Image Recognition and Microtask Crowdsourcing

三原 鉄也\*<sup>1</sup>  
Tetsuya Mihara

石川 夏樹\*<sup>2</sup>  
Natsuki Ishikawa

豊田 将平\*<sup>3</sup>  
Shohei Toyota

永森 光晴\*<sup>1</sup>  
Mitsuharu Nagamori

杉本 重雄\*<sup>1</sup>  
Shigeo Sugimoto

\*<sup>1</sup> 筑波大学図書館情報メディア系

\*<sup>2</sup> 筑波大学情報メディア創成学類

Faculty of Library, Information and Media Studies Univ. of Tsukuba

College of Media Arts, Science and Technology Univ. of Tsukuba

\*<sup>3</sup> 筑波大学図書館情報メディア研究科

Graduate School of Library, Information and Media Studies Univ. of Tsukuba

Frames are the most fundamental objects in the expression of manga. Many technologies have been developed to identify frame areas from manga graphics, however, there is no universal method that can identify frames as well as human readers can from various types of manga expressions. We propose a hybrid identification method of frames that combines image recognition with microtask crowdsourcing. This hybrid method brings together the accuracy of human readers, and the efficiency of dispersed identification tasks.

### 1. はじめに

マンガのコマは漫画表現の最も基本的な構成要素といえる。Will Eisner がマンガ・コミックを“sequential art”と呼び、隣り合って配置されたコマに描画された絵の差異から物語上の出来事を読み取る表現であると定義し論じた[McCloud 94]ように、マンガの読解においてコマを識別することはマンガによって表された物語を読み取る際に最初に行われる認知作業である。

マンガは紙媒体での流通を前提として制作されるのが主流であり、デジタルデータであっても紙を踏襲するように画像データで提供されることが一般的である。これまでに計算機上でコマ情報の利用やコマ単位でのマンガの再構成を目的として、マンガ・コミック画像からコマ領域を認識する技術の研究開発は数多く行われているが、一般的な手法は未だ普及していない。

その主な要因として、Ground Truthとなるマンガ画像とコマ領域のデータの不足により、それらの客観的な評価が困難な点が挙げられる。研究で利用可能なマンガの画像データセットはなど近年整備されているものの、コマ領域のデータの提供は十分な量がなく限定的である。

コマの識別は作者の表現と読者の感性に基づく読み取りに依るものであり、必ずしも描画の特徴からコマを識別できるとは限らず、その機械的な認識は困難である。他方、マンガは1冊あたりおよそ 100~200 ページあり、機械的認識手法の評価や機械学習のために必要な量のコマ領域を手で全て指定する手間は膨大である。こうした問題がマンガのコマ認識の一般的な手法の確立を阻害する要因になっていると考えられる。

そこで本研究ではマンガ画像から機械的に抽出したコマ領域し、それらについてマイクロタスク型クラウドソーシングを用いて正誤判定することでコマ領域を識別する手法を提案する。この手法により、従前は困難であったマンガのコマ領域のデータ作成の正確さとスケーラビリティの両立を実現することが期待できる。本稿では提案手法及び提案手法を実現したシステムとマイクロタスクに関する評価実験について述べる。

### 2. マンガ画像からのコマ認識

本文マンガ画像からのコマ領域の認識手法には Regaud ら [Regaud 13]、野中ら[野中 12]、石井ら[石井 10]、柳沢ら[柳沢 16]などの手法がある。Regaud らは連結成分をベースとした手法を提案している。まず連結成分の検出を容易にするために入力画像をグレースケール化し、白黒の二値に変換する。二値化の閾値は境界線ピクセルの平均値を閾値とする。白と黒を反転して連結成分を抽出して背景と要素を分け、それぞれの要素のバウンディングボックスをコマやテキストに分類し、コマを検出する。野中らはマンガのコマは矩形で表現され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同定する方法を提案した。石井はコマの角を検出することでコマの検出を行う手法で、これにより矩形でない領域や四角形以外の多角形、吹き出しがはみ出している等の辺が連続でない領域に対しても検出を可能にしている。柳沢らは畳み込みニューラルネットワークを用いた一般物体検出手法である R-CNN を改良した Faster R-CNN を用いてコマを検出する方法を提案している。

これらの手法はその特性上、例えば、枠線を発見することでコマを認識する手法は枠線がないコマの発見が難しい、といった認識が得意なコマの特徴に大きなばらつきがある。評価実験の結果もそれぞれ高いものであるといえるが、評価に用いている画像が研究ごとに異なるため、実際のマンガ画像一般に対して、それらの手法の有用性が検証されているは言えない。研究で利用可能なマンガの画像データセットとして Manga109[Manga109 15]や eBDtheque[eBDtheque 13]などが提供されており、評価用データを共通に利用することは可能になっているが、コマ領域の正解データセットの提供は網羅的に行われておらず、不十分である。

そもそも、これらの研究ではコマとは何かという概念、定義が明示されておらず、認識と対象とするコマの条件の一致をみていない。コマは記号的な描画によって読者に認知されるものであり、その描画法は多様である。表現は時として敢えてルールをはみ出すことで強い印象を与えるものであり、コマについても描画状の特徴からその全てを識別することはほぼ不可能であると言える。

従ってコマの認識には何らかの方法で読者による判断を組み込むが必要になると考えられる。

連絡先: 三原鉄也, 筑波大学図書館情報メディア系, 茨城県つくば市春日 1-2, mihara@slis.tsukuba.ac.jp

3. 画像認識とマイクロタスク型クラウドソーシングを組み合わせたマンガのコマ領域の判定手法

3.1 マイクロタスク型クラウドソーシング

クラウドソーシングとは一人だけでは解決が困難な課題に対し、不特定多数の人に作業を発注し、解決を行う仕組みのことである。クラウドソーシングのうち、ごく短時間で一回の作業が可能になるようにタスクを細分化して、それを複数の作業者に提示し分担して作業する手法としてマイクロタスクがある。このマイクロタスクは、特別なスキルを必要としない作業が中心であり、幅広い個人に作業を委託することができる。

本研究での提案手法では、現状では人の認知に依ることが必須であるコマ領域の判定をマイクロタスク化しクラウドソーシングすることで、スケーラビリティを確保し、効率的なコマ領域の判定を実現する。

3.2 マンガ画像からの自動検出と正誤判定マイクロタスクの組み合わせによるコマ領域の判定

提案手法では、まず先行研究で提案された形状認識アルゴリズムを用いた画像認識によってマンガ画像ごとにコマ領域の候補を判定する。次に、その候補の正誤の判定をマイクロタスクでワーカーに依頼し、複数の回答の多数決で正誤判定を行う。またこの回答結果から認識されたコマ領域が不足するページについては、認識のためのパラメータを調整して再び画像認識を行い、検出された候補をワーカーに問い合わせる。

4. コマ領域が持つ特徴に関する基礎検討

提案手法を具体的に検討するにあたり、まず認識すべきコマの領域の定義を検討するために、いくつかの実際のマンガのコマの特徴を調査した。具体的には商業出版されたマンガ単行本 4 作品5冊について、コマの特徴を以下の3パターンに分類し、その出現割合を比較した。

- ① 明確な枠線、ページの端部によって領域が閉じられている
- ② 枠線で閉じられているが、オブジェクトが重なっており不完全
- ③ 全く枠線によって閉じられていないが、オブジェクトの集合が存在する

表 1 は比較した 4 作品5冊について、①～③の分類結果を示す。この結果から、コマ領域の特徴の出現傾向は作品によって様々であり、汎用的にコマ領域を認識するためには枠線による完全な閉領域を検出するだけでは不十分であり、オブジェクトにより不完全な閉領域であるコマを検出する必要があることが示唆された。

5. コマ領域の判定システムの実装

図1は判定処理のフローチャートである。本手法では、比較的精度の高い Rigaurd の連結成分抽出をベースとした手法に

表 1 コマの特徴の分類結果

作品	ジャンル	①のコマ(%)	②のコマ(%)	③のコマ(%)
A-1	ヒューマンドラマ	91.57	8.42	0
A-2	ヒューマンドラマ	87.12	12.87	0
B	SF	21.62	78.37	0
C	ラブコメ	61.0	34.0	5.0
D	少女マンガ	38.51	50.37	11.11

基づいた画像認識プログラムを実装し、マンガ画像からのコマの自動検出に用いた。このプログラムでは入力画像を二値化し、連結成分の特徴からコマ候補となるバウンディングボックスを検出する。さらに、コマとする最小の幅をページの幅の 10%、最小の高さをページの高さの 10%、最大の高さをページの高さの 95%とするコマ候補と不要なバウンディングボックスを分けるフィルタを適用し、コマ候補を検出する。

次にコマの自動検出結果の誤検出を除去し、検出結果の正誤判定をするマイクロタスクを筑波大学融合知能デザイン研究室が運営するクラウドソーシングプラットフォームである Crowd4U[Crowd4U 11]を用いて実装した。

まず画像認識によってコマ領域が抽出される。次にその結果を図 2 のようなタスク画面でマンガ画像上に赤く表示する。そしてワーカーに赤い枠線の内側の領域がコマとして適切であるかどうかを質問する。質問に対する回答は「はい・いいえ・わからない」で回答することができる。結果の信頼性を高めるために同じタスクを別々の 3 人に発行し、その 3 人の回答から多数決を取り正解を決定する。

さらにクラウドソーシングによって認識されたコマ領域が不十分な画像について、再認識を行う。この繰り返しを行う条件はマンガ画像に含まれる正解のコマの面積の合計がそのマンガ画像の面積の 70%に達したかどうかで判断する。70%に達しているページはそこで終了し、70%に達しなかったページはもう一度画像認識とクラウドソーシングを繰り返す。この画像認識とクラウドソーシングの組み合わせの繰り返しの上限は 3 回とする。

再認識を行う際には 1 度目に取りこぼしたコマも検出できるようにパラメータの設定を変更する。まず二値化した入力画像のうちクラウドソーシングの結果正しいと判定されたコマの領域を黒く塗りつぶす。これによって再度画像認識を行う領域を狭める。この後の処理は 1 度目の画像認識と同様にして連結成分のバウンディングボックスをコマ領域の候補として抽出する。このとき、抽出されたコマ領域の候補のうち不要な候補を除去するフィルタのパラメータを最小幅については 10%から 5%に、最小高さについてはページの高さの 10%から 5%に、最大高さについてはページの高さの 95%から 100%に緩和する。

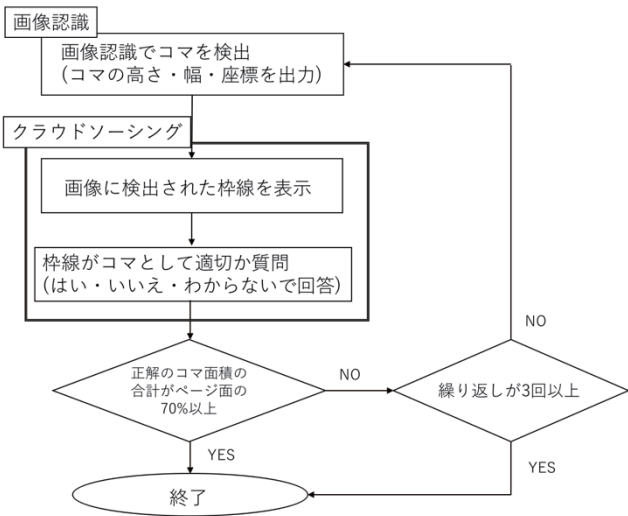


図 1 コマ領域を決定するマイクロタスクのフローチャート



図 2 コマの領域を決定するマイクロタスクのタスク

6. 評価実験

6.1 実験概要

5 章で示した 3 つのシステムの効果を実証するために、コマの領域を決定するマイクロタスクを実装し、評価実験を行った。この評価実験では、マンガメタデータのコマ情報の付与作業の効率性、及びマイクロタスクの設計の適切性を示す。またマイクロタスクを行なったワーカの回答の品質の評価と、得られた結果から 4 章で提案したコマの定義の妥当性についても検証する。

実験はマンガ作品 A[佐藤 02]1 巻・2 巻と、作品 B[木野 05] 飛ぶ東京[木野 06]を対象として行う。この実験対象に含まれるマンガ画像は 399 枚である。マイクロタスクのワーカ数は 15 人である。

6.2 結果

まず対象の 3 冊のマンガについて画像認識を行い、その結果と正解データを比較した。作品 A-1,2 の冒頭 50 ページ、作品 B 全体の 37 ページについての結果は表 2 のようになる。検出結果の精度は作品 A では 90%以上、作品 B で 61%と、2.3 節で述べた関連研究同様作品によって精度に差があることがわかる。適合率は全ての作品で 50%以上であることから、検出結果には正解が十分に含まれておりクラウドソーシングを行うにふさわしいと言える。

次にマイクロタスクの効果の実験を行なった。その際の回答

表 2 画像認識の結果

タイトル	正解のコマ数	検出数	正しい検出	再現率(%)	適合率(%)
A-1	251	396	229	90.8	57.8
A-2	214	303	200	93	66
B	183	220	112	61	51.1

表 3 ワーカの回答の誤回答数

段階	タスク数	誤回答	誤回答の起 こる確率	わからない・ スキップ	クラウドソーシングの結 果と正解データが異なる
1 度目	8534	59	0.69%	0	0
2 度目	2541	17	0.66%	14	2

表 4 終了条件の達成度

段階	全ての ページ	終了条件を満たした ページ	終了条件を満たして いないページ	終了条件を満たしたページ に含まれる未検出のコマ
1 度目	399	298	101	3
2 度目	101	54	47	0

表 5 正しい検出結果に対するワーカの誤回答

	1 回目	2 回目
正しい検出結果	1715	390
正しい結果のタスク数	5145	1170
ワーカの誤回答	38	9
誤回答率	0.74%	0.76%

表 6 間違っている検出結果に対するワーカの誤回答

	1 回目	2 回目
間違った検出結果	1127	457
間違った結果のタスク数	3381	1371
ワーカの誤回答	21	8
誤回答率	0.62%	0.58%

の品質についての評価を 1 度目のクラウドソーシング後及び 2 度目のクラウドソーシング後の 2 段階に分けて行った。

表 3 は各段階での回答の誤回答数を示している。各段階でのワーカの誤回答は 1%以下であった。これにより、今回実装したマイクロタスクがワーカにとって作業しやすいものであり、タスクの品質を保持できるものであることが示されている。

表 4 はタスクの終了条件の達成度を示している。1 度目のクラウドソーシング終了時に終了条件を満たしているページに含まれる検出されていないコマは 3 コマのみであり終了条件の設定も適切であると言える。2 度目のクラウドソーシング終了後に終了条件を満たさなかったページの特徴は 1 度目の画像認識の結果とあまり検出結果が変わらない、またはパラメータを緩和したことにより誤検出が増えているページであり、これ以上の検出結果の改善は見込めないで 3 度目は行わずに終了とする。この終了条件を満たしていないページについては手作業で修正する。終了条件を満たしていない 47 ページに含まれるコマのうち検出できておらず手作業によるコマ領域の同定作業が必要なコマは 108 コマであった。実験の対象作品に含まれる全てのコマの合計は 2027 コマであり、コマを同定する手作業の工数が 94.7%削減されたことが確認できた。

表 5、表 6 はそれぞれ抽出されたコマが正しい場合(正例)と間違っている場合(負例)のワーカの誤回答について示したものである。いずれの誤回答も明確な誤りとわかるものであったため、ワーカ間のコマの解釈の違いと言った結果に影響するものではないと考えられる。2 度目のクラウドソーシングでは、誤回答は正例・負例共に 1 度目と同様の確率で発生しているが、1 度目と異なる点は「わからない・スキップ」の回答が 14、我々の作成した正解データと一致しないものが 2 つ存在する点である。正解データと一致しなかった抽出コマの例を図 3、図 4 に示す。図 3 では実際のコマの領域に対して検出された領域が小さいがワーカは 3 人ともコマとして適していると回答している。よってワーカが求めるコマの抽出精度はこの程度であることが考えられる。図 4 では二つのコマをまとめて検出している検出結果に対してコマとして適していると回答している。これについては 2 つのコマの間の余白が狭く、オブジェクトのはみ出しもあるため境界線が



わかりにくく、タスク画面上でコマが分割されていることが認識できずに生じた誤回答だと考えられる。これを防ぐために検出領域を表す枠線を半透明、または点線にするといったユーザインタフェースの改良が必要である。このようにクラウドソーシングの結果と正解データが異なる場合であってもコマの解釈の違いから生じた違いではないと考えられ、ワーカの判断したコマ領域は全て第 4 章で述べたコマの定義に含まれているものであることが確認できた。また誤回答の発生する分布について、1 度目・2 度目ともに作品ごとの偏り、ページごとの偏りは見られず、ワーカにとって回答が難しい作品やページが存在するような傾向は見られなかった。

表 7 は手作業の場合及び各タスクの作業時間を示したものである。メタデータ作成に習熟した作業者が Comic Meta Editor[Comic Meta Editor 14]を用いて全て手作業でコマを同定するために6時間 39 分を要する一方で、クラウドソーシングを行なったのちに手作業でコマを同定する作業の時間は 2 分程度であった。これにより習熟した作業者が必要となるタスクの作業時間を 99.1%削減したと言える。また、2 回の画像認識の要した時間は 6 分ほどで、2 回のクラウドソーシングに要した時間は 6 時間ほどであり、タスク全体の作業時間を手作業の場合と比較しても 3%の作業時間を削減できた。

7. おわりに

本研究では、メタデータ作成の効率化を目指して画像認識とクラウドソーシングを組み合わせたマイクロタスクを実装した。ユーザと研究者の求める品質の差などの問題はあがるが、マンガメタデータ作成にかかる手作業の工数を大幅に削減し、計算機だけによるコマ抽出では不十分な精度を向上させることが可能



図 3 タスク結果と正解データが異なる例 1



図 4 タスク結果と正解データが異なる例 2

表 7 作業時間の比較

方法	作業対象	作業時間	総作業時間
1.手作業のみ	399ページ	60秒/ページ	6時間39分
2.画像認識1回目	399ページ	0.7秒/ページ	4分39秒
3.クラウドソーシング1回目	8534タスク	2秒/タスク	4時間44分
4.画像認識2回目	101ページ	0.7秒/ページ	1分10秒
5.クラウドソーシング2回目	2541タスク	2秒/タスク	1時間24分
6.クラウドソーシング後の手作業	未検出の111コマ	10秒/コマ	2分18秒
7.2+3+4+5+6			6時間16分7秒

になった。本手法はそれ自身がマンガのコマ領域の判定手法として有用であると同時に、特に正解データセットを作成する手法として期待される。

なお本手法は実際に crowd4 内で LODEM Project[LODEM Project 17]で公開しており、タスクを実施中である。今後はコマ領域を判定する画像の追加やそのほかのマンガの構成要素のメタデータの作成タスクを公開する予定である。

謝辞

本研究における Crowd4U の利用とマイクロタスクの実装に当たっては融合知能デザイン研究室及び同研究室森嶋厚行教授、橋本大空氏の協力を頂いた。ここに深謝する。

参考文献

[McCloud 94] Scott McCloud: Understanding Comics : The Invisible Art, William Morrow Paperbacks, 1994.

[Rigaud 13] Christophe Rigaud, Norbert Tsopze, Jean-Christophe Burie, Jean-Marc Ogier: Robust frame and text extraction from comic books. Graphics Recognition, New Trends and Challenges Lecture Notes in Computer Science, pp.129-138, Springer, 2013.

[野中 12] 野中俊一郎, 沢野拓也, 羽田典久: 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 RESEARCH & DEVELOPMENT, No.57, pp.46-49, 富士フイルム株式会社, 2012.

[石井 10] 石井大祐, 河村圭, 帆足啓一郎, 滝嶋康弘, 渡辺裕: コミック画像におけるコマの角検出に関する一検討, 情報処理学会研究報告, 2010.

[柳沢 16] 柳沢秀彰, 渡辺裕: Faster R-CNN を用いたマンガ画像からのメタデータ抽出, 2016 年映像情報メディア学会年次大会, 2016.

[佐藤 02] 佐藤秀峰: ブラックジャックによろしく, 講談社, 2002.

[木野 06] 木野陽: 飛ぶ東京, 自費出版, 2006.

[Manga109 15] <http://www.manga109.org/ja/>, 2015.

[eBDtheque 13] <http://ebdtheque.univ-lr.fr/>, 2013

[Comic Meta Editor 14] ComicMetaEditor2014, <https://github.com/mangaresearch/ComicMetaEditor2014>, 2014.

[Crowd4U 11] <http://crowd4u.org/ja/>, 2011.

[LODEM Project 17] <http://crowd4u.org/ja/projects/lodem>, 2017.