

マンガの構成要素に基づく 自動シーン分割処理に関する一検討

石井 大祐^{1,a)} 柳澤 秀彰² 三原 鉄也³ 永森 光晴⁴ 渡辺 裕²

概要：物語性のあるマンガにおいて、シーン情報は重要なメタデータの一つである。シーンはマンガの意味的側面における一区間として定義されており、内容の要約単位として有効である。また、マンガ内に含まれる多数のメタデータをシーン単位にまとめることで、メタデータの利便性を高められるものと考えられる。これまでに、マンガ画像解析処理として、コマ、登場人物、吹き出し等、マンガの構成要素を検出するための手法が提案されている。我々の研究目的は、マンガのコマ内に含まれる構成要素に基づいて、シーンの切り替わりとなるコマを判別する手法を実現することである。本稿では、マンガの各コマに含まれる構成要素の分布について調査し、シーンの切り替わりとなるコマについて重要な構成要素を明らかにする。実際のマンガ画像から取得したメタデータを基に調査を行った。調査結果から、他のメタデータと比較して、現時点で自動取得可能とされるメタデータではナレーション、現時点で自動取得困難なメタデータでは背景の距離がシーン切り替わりにおいて重要な要素となりうる可能性が高いことが確認された。

A Study on Scene Segmentation Method Based on Comic Component

1. まえがき

近年マンガを対象とした画像処理、情報処理に関する研究が活発化している。マンガ関連研究は、マンガの内容理解だけでなく、表現分野や、マンガ作成支援、マンガの構造解析など幅広い分野に広がる。マンガは基本的に絵と文字からなる複合表現である。

マンガには、一つながりの話を4コマにまとめた4コママンガと、シーンが不定数のコマからなり、多数のシーンを時系列的に並べてゆくことでストーリーを構成する長編マンガがある。一般的に、長編マンガでは、一話が複数のシーンから成り立っている。長編マンガにおけるシーン

は、作者によって定義される、ストーリー上の意味のあるいは時間的な区切りである。一話のページ数は、連載作品の場合には掲載誌上の都合により決定される場合が多く、市販される単行本には複数話が収められる。本稿では4コマ漫画を除外し、長編マンガを対象とする。

シーンは、長編マンガにおける意味のあるいは時間的な区切りという側面から、自動要約における一コマ以上一話未満の情報区分単位として有用であると考えられる。マンガでは、コマ、登場人物、セリフ、ナレーション、オノマトペ、背景等多数の要素が紙面上に重畳される。これらの要素から得られるメタデータの集合単位としてシーンを用いることで、メタデータをマンガの意味的側面に基づいた形で扱うことが可能となる。

多量のメディアコンテンツを取り扱う際に、メディアコンテンツの要約が利用される。例えば、映像作品の自動要約に関する研究が行われている。自動要約においては、要約後のコンテンツにおいて、要約元のコンテンツの内容がある程度理解可能である必要がある。

マンガの自動要約を行う場合においても、抽出された結果が一コマのみではストーリーとしての情報が失われてしまい、一話では要約の一単位として大きすぎるという問題がある。一方で、シーンは意味的な連続性のあるコマの集

¹ 早稲田大学 大学院国際情報通信研究科
Graduate School of Global Information and Telecommunication Studies, WASEDA University

² 早稲田大学 大学院基幹理工学研究科
Graduate School of Fundamental Science and Engineering, WASEDA University

³ 筑波大学 図書館情報メディア研究科
Graduate School of Library, Information and Media Studies, University of Tsukuba

⁴ 筑波大学 図書館情報メディア系
Faculty of library, Information and Media Studies, University of Tsukuba

a) dai.ishii@aoni.waseda.jp

合であるため、ストーリーとしての情報を保ちつつ、要約の一単位における分量を小さくすること可能となる。

本稿では、シーンの切り替わりとなるコマの自動検出を実現するため、マンガのコマ内に含まれる構成要素の利用を試みる。本稿では、マンガの各コマに含まれる構成要素の分布について調査し、シーンの切り替わりとなるコマについて重要な構成要素を明らかにする。実際のマンガ画像から取得したメタデータを基に調査を行う。

以下2では、これまでに行われてきたマンガ画像からの自動メタデータ検出処理の概要と、現時点で自動取得可能なメタデータについてまとめる。3では実際にマンガから構成要素に関する情報を取得し、その分布について示す。4ではシーンの分割点となるコマの情報を基に解析処理を行った結果についてまとめる。5は本稿のまとめである。

2. マンガ内容に関する情報抽出処理

マンガは絵と文字からなる複合的な画像である。また、マンガ上に存在する輝度成分は基本的に0と255からなる白黒であるため、特徴抽出が困難である。これまでに、マンガ上に存在するメタデータを取得するための手法が提案されている。取得されるメタデータのターゲットはコマ、セリフ、登場人物などがある。

マンガからのコマ並びにセリフ情報の抽出処理として、富士フィルムのGT-Scan, GT-Ballonといったソフトウェアがある[1][2]。GT-Scanは、従来検討されてきた再帰的な2分割処理に基づいたマンガのコマ分割処理[3][4]と比較して高い精度でコマ検出処理が可能である。GT-Ballonは吹き出しの自動検出処理である。吹き出しの検出において、自動的に吹き出し領域を塗りつぶすことが可能とされる。

マンガの吹き出しには用途に応じたバリエーションが存在する。通常のセリフとしての意味以外に、よく使用される用途として、ナレーションがある。ナレーションはシーンの状況を登場人物あるいは第三者の視点から説明するものである。ナレーションでは、多くの場合吹き出しの形状がセリフと異なる。具体的には、一般的な吹き出し形状が楕円形である場合、ナレーションの場合には矩形となる等、読者に対して差が明となるように工夫される。上記吹き出し検出処理においては、吹き出しの形状も取得可能となるため、セリフとナレーションについては分離可能である。

マンガ画像からの登場人物の検出処理およびその識別処理[5][6]が提案されている。今日一般的に利用されている、自然画像から人間の顔を検出する処理では、Haar-Like特徴量を利用した処理が行われている。一方でマンガにおける登場人物の検出処理では、エッジの形状を捉える画像特徴量が利用される。特に、Deformable Part Modelを用いた手法[7]は従来よりも高い精度でマンガにおける登場人物の顔が検出可能となっている。登場人物間の識別処理[6]



図1 シーンの区切りとなるコマの例 文献[8](p.16, p.57)から抜粋し著者が一部編集したもの

Fig. 1 Samples of scene division frame, This image is extracted from [8] p.16 and p.57 and edited partly by author.

も同様の処理により行われる。現状では、学習用画像を集めやすい登場回数が多い人物について、未知画像の正解画像に対し0.79の識別成功率が得られている。

3. マンガの構成要素とその分布

今回取り扱うメタデータを以下に示す。2にて挙げたマンガ解析処理によって自動的に得ることが可能あるいは可能となりうるメタデータとしてコマ、セリフ、ナレーション、人物の位置、特定の登場人物の有無、文字情報の長さ、コマの大きさが挙げられる。また、現状では取得困難なメタデータとして、オノマトペ、話者に対応のとれたセリフ情報、背景の距離などがある。本稿では、これらの構成要素情報から、シーンの分割点を自動的に検出するために重要な要素について検討する。

データを取得した作品数は1作品とした[8]。今回扱うシーンの区切り箇所については、作者の協力により得たGround Truthを使用する。シーンの区切りとなるコマの例を図1に示す。図中にて点線で示されたコマがシーンの切り替え点となるコマである。空白やタイトルを除いた総ページ数は191ページ、合計コマ数は956、1ページに含まれる平均コマ数は5.01であった。この中で、シーンの切り替わりとなるコマ数は63である。各要素値はコマ毎に、以下のルールに基づいて算出した。

Scene Change

シーンの切り替え点となる場合に1、切り替え点ではない場合に0とする

Character

登場人物の総数

Dialog, Narration

コマに属する吹き出しの総数

Onomatopoeia

コマに属するオノマトペの総数

表 1 マンガ構成要素
Table 1 Contents of comics

	Total	Average	Median	Variance
Scene Change	63	0.066	0	0.062
Character	1448	1.515	1	1.425
Dialog,Narration	1595	1.668	2	1.491
Onomatopoeia	209	0.219	0	0.213
Main Character	1058	1.107	1	0.629
Other Character	380	0.397	0	1.012
Narration	58	0.059	0	0.116
Speech(Main)	1242	1.299	1	1.422
Speech(Other)	287	0.300	0	0.673
Long Dialog	176	0.184	0	0.232
Short Dialog	1409	1.474	1	1.328
Background Distance	439	0.459	0	0.387
Frame Size	2811	2.940	3	0.728

表 2 シーンの区切りとなるコマのマンガ構成要素
Table 2 Contents of comics on frame of scene border

	Total	Average	Median	Variance
Character	85	1.349	1	4.392
Dialog,Narration	55	0.873	0	1.338
Onomatopoeia	14	0.222	0	0.240
Main Character	42	0.667	0	0.645
Other Character	43	0.682	0	3.543
Narration	4	0.063	0	0.060
Speech(Main)	41	0.651	0	0.941
Speech(Other)	10	1.587	0	0.458
Long Dialog	6	0.095	0	0.120
Short Dialog	48	0.762	0	0.959
Background Distance	61	0.968	1	0.515
Frame Size	176	2.794	3	1.166

Main Character

作品の中で重要な役割を担う登場人物の総数

Other Character

非主要登場人物数

Narration

Dialog,Narration の内のナレーション数

Speech(Main)

Dialog,Narration の内、主要登場人物のセリフ数

Speech(Other)

Dialog,Narration の内、非主要登場人物のセリフ数

Long Dialog

Dialog,Narration の内、4 行以上である数

Short Dialog

Dialog,Narration の内、4 行未満である数

Background Distance

なし=0, 近距離=1, 遠距離=2 とした評価値

Frame Size

小さい=1, 普通=2, 大きい=3, 1 ページ=4

ただし、Background Distance については、1 コマに 2 種類以上の背景距離が存在する場合、最も遠いものをそのコマの背景距離として採用することとした。

実際上記メタデータを人的に取得した結果を表 1 にまとめる。マンガ構成要素の分布を図 2 に示す。また、シーンの区切りとなるコマのマンガ構成要素を図 2 に示す。ここで、Total は本作品における各要素値の総合計であり、Average は一コマ単位での平均要素値、Median は再頻出値、Variance は分散である。

4. シーン分割点に関する解析結果

前節にて得られた構成要素情報を基に、マンガのシーンの分割点となるコマについて影響の高い構成要素を検討する。ここでは、現時点で自動取得可能なマンガの構成要素として、Character, Dialog, Main Character, Narration,

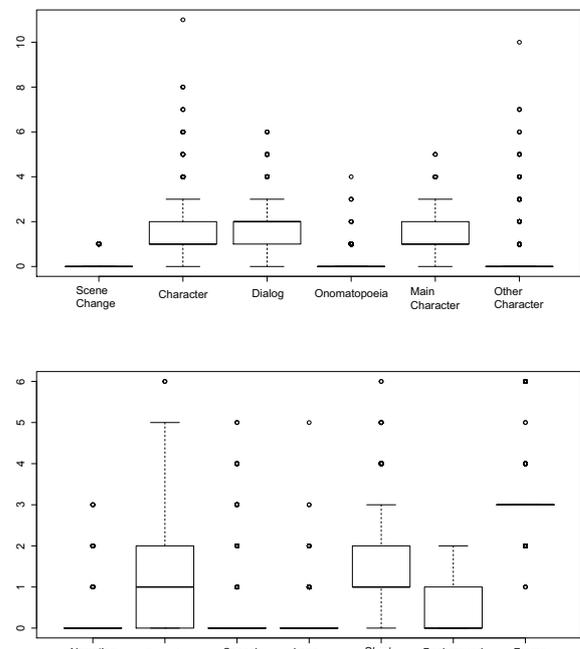


図 2 マンガ構成要素の分布

Fig. 2 Distribution of Comic Contents

Speech, Frame Size の 6 種類を用いた場合 (Detectable) と、前節で得られた構成要素情報全てを用いた場合 (All) について調査を行った。

今回判別対象となる要素はシーンの分割点であり、値としては該当するもしくは該当しないの 2 値となる。そこで、ロジスティック回帰分析により、各要素の影響度合いを比較する。利用した構成要素ごとのロジスティック回帰分析結果並びにオッズを表 3 および表 4 に示す。ここで、Detectable の場合においては、その取得精度が低い場合を考慮し、各要素の数ではなく、各要素のコマにおける存在有無を 1 もしくは 0 の 2 値データとして入力する。また、Speech は Speech(Main) と Speech(Other) のどちらか

表 3 ロジスティック回帰分析結果 (Detectable)
Table 3 Result of Logistic regression (Detectable)

	Estimage	Std. Error	P-Value	Odds
Character	-1.492	0.527	0.005	0.225
Dialog	-3.063	0.102	0.002	0.047
Main Character	-0.219	0.507	0.666	0.803
Narration	2.245	0.841	0.008	9.438
Speech	1.654	1.005	0.100	5.228
Frame Size	-0.189	0.140	0.178	0.828

表 4 ロジスティック回帰分析結果 (全体)
Table 4 Result of Logistic regression(ALL)

	Estimage	Std. Error	P-Value	Odds
Character	-0.900	0.941	0.339	0.407
Dialog	-0.420	1.362	0.758	0.657
Onomatopoeia	-0.415	0.306	0.175	0.660
Main Character	0.225	0.955	0.814	1.253
Other Character	1.053	0.944	0.264	2.867
Narration	0.380	1.249	0.761	1.462
Speech(Main)	0.643	1.178	0.585	1.902
Speech(Other)	0.359	1.163	0.758	1.432
LongDialog	-0.623	0.934	0.505	0.536
ShortDialog	-0.856	0.831	0.303	0.425
Background Dist.	1.170	0.204	0.000	3.222
Frame Size	-0.295	0.162	0.070	0.745

表 5 SVM による識別処理結果
Table 5 Discrimination results by SVM

	No	Yes
No	425	0
Yes	31	0

が存在する場合には 1, どちらも存在しない場合を 0 とした。All の場合には前節で得られたデータを直接入力する。

ロジスティック回帰分析の結果から, 本作品では, 現時点で自動取得可能な情報を用いた場合においては, ナレーションの有無, 全データを使用する場合には背景の距離が, シーンの切り替えとなるコマに対し最も影響力が大きいことが確認された。

さらに, 全データを用いた場合の自動判別性能を調べるため, SVM による学習と判別処理性能について調査を行った。学習には 500 コマを使用し, 学習後の誤り率は 0.046 となった。本学習結果を用いて学習に用いていない 456 コマの解析を行った。識別結果を表 5 に示す。ここで, 横軸が実際の回答, 縦軸が正解を示す。表より, 全てのコマをシーンの分割点ではないと判定する結果が得られた。この原因として, シーンの分割点であるコマ自体に含まれる要素の分散が大きいことが考えられる。

5. むすび

本稿では, マンガのシーンの区切りとなるコマを識別す

るため, コマに含まれる要素について調査を行った。マンガの各コマに含まれる構成要素の分布に対する解析結果より, 現時点で自動取得可能とされるメタデータでは, ナレーション, 現時点で自動取得困難なメタデータでは背景の距離がシーン切り替わりにおいて最も影響力が大きい要素であることが確認された。一方で, SVM によるシーン区切りの自動識別では, 今回利用した情報のみからでは識別が不可能であった。これは, 今回の調査で利用したマンガにおいては, シーンの区切りとなるコマに含まれる要素の分散が非常に大きいため, 学習により十分な識別性能が得られなかったことが原因であると考えられる。マンガのシーンの区切りにおいては, 区切りとなるコマのみならず, その前後のコマと連動して特徴が得られる可能性も考えられるため, 識別器を構成する際に前後のコマに含まれる要素の分布を加えることで, シーン区切りの自動識別が可能になる可能性が考えられる。

謝辞 本研究成果の一部は JSPS 科研費 26870637 の助成を受けたものである。本稿において図 1 にて例として示し, また, 実験に用いた画像 [8] について, 利用を許可していただいた学研教育出版様に深く感謝する。

参考文献

- [1] 野中俊一郎, 沢野哲也, 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, 富士フィルム研究報告, Vol.57, pp.46-49, 2012.
- [2] 野中俊一郎, 寺横 素, 沢野哲也, 羽田典久: 電子コミック向け画像処理技術の開発とその応用, 富士フィルム研究報告, Vol.58, pp.42-47, 2013.
- [3] T.Tanaka, Kenji Shoji, and J. Miyamichi: Layout analysis of tree-structured scene frames in comic images, Proc. IJCAI-07, pp.2885-2890, January 2007.
- [4] 石井大祐, 河村圭, 渡辺裕: コミックのコマ分割処理に関する一検討, 電子情報通信学会論文誌 D, Vol.J90-D, No.7, pp.1667-1670, July 2007.
- [5] 新井俊宏, 松井佑介, 相澤清晴: マンガ画像からの顔検出, 電子情報通信学会総合大会, D-12-67, p.161, March 2012.
- [6] 石井大祐, 渡辺裕: 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, 2013.
- [7] 柳澤秀彰, 石井大祐, 陳明, 渡辺裕: 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, 映像情報メディア学会年次大会, 17-9, September 2014.
- [8] 布袋あずき, ミハラテツヤ: マンガジュニア名作シリーズ 小公女, 学研教育出版, 2012.