

フィルムコミックの自動生成における視線情報の利用

澤田 友哉[†] 豊浦 正広[‡] 茅 暁陽[‡]

[†] 山梨大学工学部コンピュータメディア工学科 〒400-8511 山梨県甲府市武田 4-3-11

[‡] 山梨大学大学院医学工学総合研究部 〒400-8511 山梨県甲府市武田 4-3-11

E-mail: [†] t08kg018@gmail.com, [‡] {mtoyoura, mao}@yamanashi.ac.jp

あらまし フィルムコミックとは、アニメーション作品をコミック調画像に変換したものである。従来、フィルムコミックの自動生成のためには、画像特徴の抽出によってショットを代表する重要フレームを選択し、同じく画像特徴の抽出によって重要箇所を削除・隠蔽しないような画像のトリミングや吹き出しの配置が行われてきた。しかし、重要フレームや重要箇所の検出には、映像作品そのものの理解が不可欠であり、画像特徴のみからこれを実現することは難しい。そこで本研究では、鑑賞者の視線情報から重要フレームや重要箇所の抽出を行い、フィルムコミックの自動生成に利用する手法を提案する。主観評価実験により、ストーリーの表現に重要なフレームの選択、トリミングにおける原構図の保存、重要箇所を隠さない吹き出しの配置がそれぞれ実現できることを確かめた。

キーワード フィルムコミック, 視線情報, Attention Map

Automatic Creation of Film Comic Using Eye-Tracking Data

Tomoya SAWADA[†] Masahiro TOYOURA[‡] and Xiaoyang MAO[‡]

[†] Computer Science and Media Engineering, Faculty of Engineering, University of Yamanashi
4-3-11 Takeda, Kofu, Yamanashi, 400-8511 Japan

[‡] Interdisciplinary Graduate School of Medical and Engineering, University of Yamanashi
4-3-11 Takeda, Kofu, Yamanashi, 400-8511 Japan

E-mail: [†] t08kg018@gmail.com, [‡] {mtoyoura, mao}@yamanashi.ac.jp

Abstract Film comic is created from animation movies by selecting important frames, trimming the frame images to fit into the panels of comic and placing speech balloons at appropriate positions on the panels. Conventional technologies aiming at automatically completing these tasks based on low level image features only fail to produce good results, because the detection of important frames and important areas in each frame requires the understanding of movie and image contents as well. We propose to detect important frames and their important areas based on viewers' eye-tracking data for automatic creation of film comic. The results of subject studies support the proposed method in selecting important frames for well conveying the story, holding the structure of original frame images, and arranging balloons without occluding important areas.

Keyword Film comic, eye-tracking data, attention map

1. はじめに

アニメーション作品をコミック調画像に変換したものはフィルムコミックと呼ばれ、有名な作品の多くについて一般に販売されている。フィルムコミックは、作品内容を一覧表現する資料として、また、映像とは別の形態をとる作品として利用される。フィルムコミックの作成は、専門家により手作業で行われてきた。図1に示すように、映像をショットに分割し、各ショットからストーリーを表現する上で必要なフレーム、および、セリフに対応するフレームを抜き出し、コミックのコマに整形して作成される。セリフはコマ中の吹き出しとして表現される。

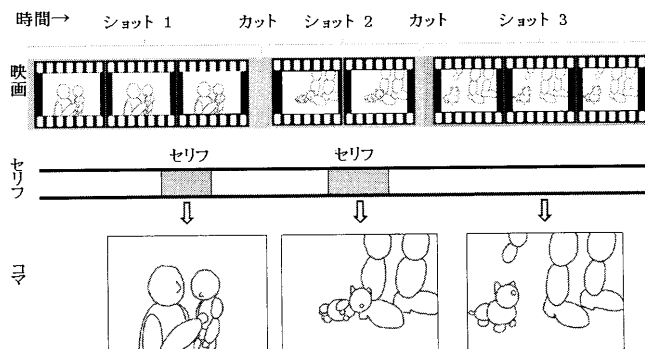


図1. 映画からコミックへの変換方法

映画をコミックに自動変換する既存技術として Hong ら[1]の手法がある。彼らの手法では、既存の顔認識及び唇の動き検出技術を用いてセリフと話者との対応付けを行い、吹き出しの配置を行う。しかし、顔認識及び唇の動き検出は主に実写映像用に開発された技術であるため、フィルムコミックの題材として特に多く利用されているアニメーション映像には適用できない場合が多い。

それに対し、国広ら[2]はフレームの選択には画像処理技術、トリミングと吹き出しの配置には、鑑賞者の視線情報を用いる方法を提案した。色の変化により映像をショットに分割し、ショット中央の画像を代表フレームとして選択する。そして、注視箇所をフレーム内の重要箇所と見なし、それが切り取られたり、隠されたりしないようにした。しかし、国広ら[2]の手法には(1)選出したフレームは必ずしもストーリーを理解する上で重要なフレームではない、(2)トリミングが元のフレーム画像の構図を反映していない、(3)吹き出しが重要箇所にかかってしまう、といった問題があった。これらの問題の解決は、映像内容の理解なしでは困難である。

2. 提案手法

本研究では、鑑賞者の視線情報と画像処理技術を併用することでこれらの問題の解決を試みる。鑑賞者の視線は、興味がある対象に向けられるため、重要フレームや重要領域、発話主体に関する情報を含んでいる。視線の移動を解析することで、ストーリーの展開を表現するのに必要なフレームを検出することができる。また、画像内の意味のある領域は通常何らかの画像特徴を含んでいるため、視線が向けられている領域にこのような画像特徴があるかどうかを調べることで、検出の精度の向上を図る。提案手法はさらに、視線情報と画像特徴から算出したフレーム画像の重要度マップに基づき、構図を保持するトリミングと重要箇所を隠蔽しない吹き出しの配置を、最適化手法を用いて行う。

2.1 フレーム選択

既存研究[2]では、映像中のカット検出を行い、ショットの代表フレームを選出していた。しかし、動作主体の変更や内容の変化は文脈依存であるため、これを画像処理のみで検知することは困難であった。そこで、本研究では内容に変化を含むシーンには、視線移動が生じる点に着目し、視線の大きな移動から、内容変化や動作主体の変更を検出する。これにより、同一ショットから内容の異なる複数のフレームを選択することができる。ただし、このときの視線移動が本当に

鑑賞者の注意のシフトを表しているかを判別する必要がある。我々は図2に示すように、まずフレーム画像から局所勾配特徴であるKLT[3]を抽出し、k-means法でクラスタリングを行うことで、画像中の意味のある領域を求める。そして、視線に大きな移動が生じた際、注視点がいずれかの領域内部にある場合のみ、その対象に注意がシフトしたと判定する。これにより、ストーリーを理解する上で、重要な箇所を検出できる。



(a) 意味のある視線移動 (b) 意味のない視線移動
図2. 視線における注意のシフト

提案手法では、注意のシフトが起きているフレームで、同一ショットを内容の異なる複数のグループに分け、それぞれのグループの中央のフレームを選択する。さらに、一人ではなく、複数人の視線データを用いて、より多くの人々が注意をシフトしたフレームを選択することで、さらなる精度の向上に成功した。

2.2 トリミングにおける構図の保存

選択されたフレームの形状とコマの形状が異なる場合は、コマの形状に合うようにフレーム画像をトリミングする必要がある。フレーム画像には製作者の意図した場面の見せ方を反映した構図がある。視線が最も集中している箇所を画面の中央にくるようにトリミングを行うと、図3に示すように本来の構図を壊してしまう恐れがある。そこで、提案手法では重要箇所を切り取らず、フレーム画像における重要箇所の相対位置がなるべく保存されるようにトリミングを行う。

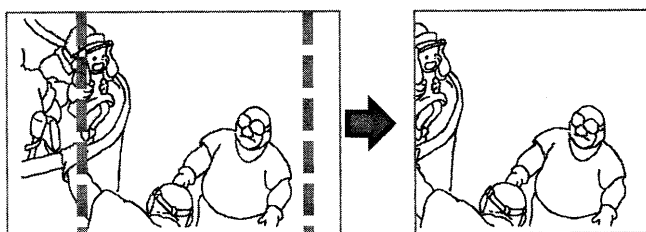


図3. 構図が保存されていないトリミングの例

重要箇所の抽出には、画像特徴と視線情報を併用する。図4に示すように、選択されたフレーム画像に対して、前後の数フレームも含めて注視位置を求めることで、内容的に重要な箇所を抽出することができる。

また、KLTを抽出することで画像特徴の観点から意味のある箇所を抽出する。それぞれの特徴から求めた重要箇所の周辺にガウス分布を利用して重みを付与させ、Attention Mapを作成する。ある位置におけるAttention Mapの値は、その位置をトリミングによって切り取ったり、吹き出しを配置して隠したりすることによるコストを与えることになる。

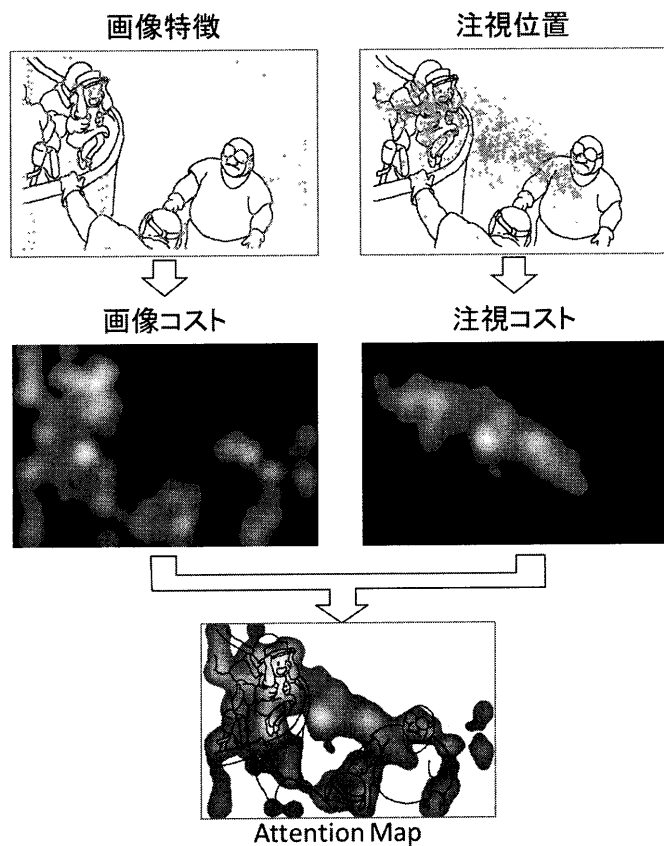


図 4. 画像特徴と注視位置からの Attention Map の作成

トリミングにより切り取られてしまうコスト E_{elim} を式(1)により計算する。

$$E_{elim} = \sum_{(i,j) \in I_{trim}} a_{ij} \quad (1)$$

ここで、 a_{ij} は切り取られる画素での Attention Map の値である。 I_{trim} は図 5 に示すように、切り取られる領域を示す。

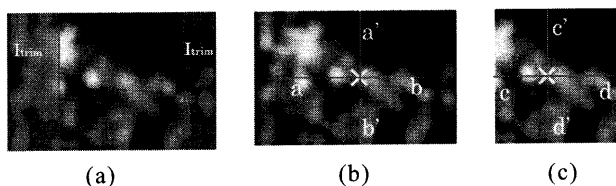


図 5. トリミングにおけるコストの計算((a) E_{trim} のコスト, (b)切り取る前のある画素の比率, (c)切り取った後のある画素の比率)

また、重要な箇所の画像上での位置ができるだけ保存されるように、式(2)のように構図のコスト E_{str} を定義する。図 5 に示すように、 E_{str} が小さいほど、重要箇所の画像上での位置が変化しないことになる。

$$E_{str} = \sum_i \sum_j a_{ij} \times \left| \left(\frac{a}{b} - \frac{c}{d} \right) + \left(\frac{a'}{b'} - \frac{c'}{d'} \right) \right| \quad (2)$$

これらのコストの和が最小となるような切り取り位置を求めることで、重要箇所の相対位置が保存されるようなトリミングを行うことができる。

2.3 吹き出しの配置

セリフのあるコマでは、コマ内の重要箇所を隠さないように吹き出しを配置する必要がある。国広ら[2]は、視線が最も集中している箇所を避けて吹き出しを配置した。しかし、注視されていない領域でも会話の内容に関連する有用な情報が含まれている場合がある。図 6 に示す白雪姫の例では、視聴者の視線を集めた会話の焦点となっている小鳥を避けて配置した吹き出しが、話者の顔の一部を覆い隠してしまっている。

本研究では、2.2 で求めた Attention Map を利用して、重要でない箇所に吹き出しを配置することでこの問題を解決する。吹き出しによって覆い隠される画素の Attention Map 値の合計を式(3)のコスト E_{hide} とし、それを最小にすることで吹き出しの位置を得る。しかし、重要度の低い箇所は、ほとんどの場合は画像の端になる。話者や注目点になるべく近い位置に配置できるように、式(4)のように吹き出しの重心から、Attention Map においてもっとも重要度の高い位置までの距離 E_{dist} もコストに加える。さらに、形状が極端に細長くならないように、コストに式(5)で示す吹き出しのアスペクト比 E_{shape} の項も含める。式(5)において、 L_{width} は吹き出し形状の横の長さを、 L_{height} は吹き出し形状の縦の長さをそれぞれ表す。

$$E_{hide} = \sum_{(i,j) \in I_{balloon}} a_{ij} \quad (3)$$

$$E_{dist} = \sqrt{(x_a - x_g)^2 + (y_a - y_g)^2} \quad (4)$$

$$E_{shape} = \left| \frac{L_{width}}{L_{height}} \right|^2 \quad (5)$$

これらのコストの和が最小になるような位置を求めることで、重要箇所を隠さず、発話対象と思われる箇所になるべく近い位置に、吹き出しを配置できる。



図 6. 適切でない吹き出し配置

3. 実験

実験では、解像度が 1920×1200 である 24 インチ液晶ディスプレイに映像を提示し、視線追跡装置 NAC 社製 EMR-AT VOXER を用いて、60Hz で視線追跡を行った。鑑賞者は、20 代の男女 5 名であり、実験に用いた映像は「紅の豚」の冒頭 15 分間とした。

提案手法を用いて作成したフィルムコミックを、これとは別の 20 代の男女 6 名に評価してもらった。

まず、フレーム選択の効果を調べた。実際に販売されている本物のフィルムコミック、既存手法で作成したフィルムコミック、提案手法で作成したフィルムコミックについて、同一のシーンを指定し、使用されている画像のみを提示した。被験者は画像のみの情報から、その間の動作や情景を書きとる。これに対し、動画を見た第三者が映画の内容を最大として 5 段階で判定を行った。分散分析を行ったところ、有意水準 1% で有意に違いがあったため、多重比較検定を行った。その結果、提案手法と既存手法、本物のフィルムコミックと既存手法の間で、有意水準 1% で違いがあり、提案手法と本物のフィルムコミックの間に有意差は見られなかった。よって、提案手法は本物のフィルムコミックに近いフレーム選択の仕方をしているといえる。

次に、トリミングの効果を調べた。3 つのシーンについて、実際の映画の画像と、それに対して既存手法と提案手法でトリミングを行った画像を提示し、(a) 重要箇所を切り取っていないか、(b) 構図を保っているかを 5 段階で評価してもらった。分散分析結果を表 1 に示す。実験結果により、提案手法は既存手法に比べて重要箇所を切らずにトリミングできており、かつ適度に構図を保持していると考えられる。

最後に、吹き出しの配置について調べた。3 つのシーンに対して、(c) 重要箇所を隠していないか、(d) 話者に近いかを 5 段階で評価してもらった。分散分析の結果を表 2 に示す。表 2 の結果より、重要箇所にかかることを抑えつつ配置を行えていることがわかる。しかし、話者までの距離については、既存手法と提案手法の間に有意な差は見られなかった。これは、既存手法が重要箇所にかかってしまおうとも、発話主体に近づ

けているためであると考えられる。

表 1. トリミングに対する分散分析の結果

(+: $p < .10$, *: $p < .05$, ****: $p < .001$)

シーン	質問(a) 重要箇所を切り取っていないか	質問(b) 構図を保っているか
1	$p = 0.0172^+$	$p = 0.5301$
2	$p = 0.0925^+$	$p = 0.0301^*$
3	$p = 0.0000^{****}$	$p = 0.0002^{****}$

表 2. 吹き出し配置に対する分散分析の結果

(***: $p < .05$, ****: $p < .001$)

シーン	質問(c) 重要箇所を隠していないか	質問(d) 話者に近い
4	$p = 0.0030^{***}$	$p = 0.1438$
5	$p = 0.1747$	$p = 0.8560$
6	$p = 0.0009^{****}$	$p = 0.4017$

4. まとめ

本研究では、視線情報を用いたフィルムコミックの自動生成手法を提案した。鑑賞者の視線移動と画像特徴のある領域を調べることで、ストーリーの表現に重要なフレームの選択を可能にした。視線位置と画像特徴から画像上の重要箇所を示す Attention Map を作成することで、画像のトリミングにおける原構図の保存と、重要箇所を隠さない吹き出しの配置をそれぞれ実現した。

今後の課題として、音声情報や生体信号の利用を挙げる。音声情報からは内容が盛り上がる時刻を知ることができ、また、効果音の表現をフィルムコミック中に挿入することもできる。鑑賞者の生体信号からは、感動が起こった時刻や内容が切り替わる時刻を調べることができると考えている。

文 献

- [1] R.Hong, X.Yuan, M.Xu, M.Wang, S.Yan, T.S.Chua, "Movie2Comics: a feast of multimedia artwork," ACM International on Conference Multimedia, pp.611-614, 2010.
- [2] 国広 守, 茅 暁陽, "視線情報を用いた動画からのフィルムコミックの自動生成," Visual Computing / グラフィクスと CAD 合同シンポジウム, Article 13, 2008.
- [3] J.Shi, C.Tomasi, "Good features to track," IEEE Conference on Computer Vision and Pattern Recognition, pp.593-600, 1994.