IIEEJ 論文

iMapを利用したフィルムコミックの自動生成

澤田友哉† 豊浦正広†† 茅 暁陽††(正会員)

†山梨大学大学院医学工学総合教育部，††山梨大学大学院医学工学総合研究部

Film Comic Generation with iMap

Tomoya SAWADA†, Masahiro TOYOURA††, Xiaoyang MAO††(*Member*)

†Department of Education Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi,
††Faculty of Engineering Interdisciplinary Graduate School of Medical and Engineering, University of Yamanashi

〈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，鑑賞者の視線情報と画像特徴からiMapを作成し，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．

キーワード：フィルムコミック，視線情報， Visual Attention，iMap， 動画処理

<**Summary**> 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.

Keywords: Film comic, eye-tracking data, visual attention, iMap, video summarization

1. はじめに

主にCGやアニメーション映画において，フィルム画像とセリフをコミックのコマや吹き出しなどで表現したものをフィルムコミックという．フィルムコミックは，日本国内のみならず海外でも人気が高く，有名なアニメーション映画について製作されることが多い．フィルムコミックは，映像と活字の楽しさを併せ持つような編集が可能である．つまり，コマ画像の間のストーリーや音声，効果音を読者が想像し，セリフを読み進めていくことができるため，一般的な活字媒体に対して画像がある分，理解が容易となる．このことは，一般的なマンガについてもいえる．読者層も広く，最近ではフィルムコミックの電子化も行われてきている．従来，フィルムコミックの作成は専門家が手作業で行ってきた．つまり，映像から使用するフレームを手動で選び出し，コミックのコマのレイアウトに従って，選んだフレームのトリミングを行い，映画のセリフ情報を吹き出しにして配置するのである．映画は膨大な量のフレームから成るため，これらの処理には非常に長く時間がかかる．また，使用するフレーム画像や配置する画像の構図など，特殊なセンスを要する作業工程が存在する．

この問題を解決するために，近年の研究でフィルムコミックの自動生成を行うものが提案されている[1)〜5)]．Hongら[4)]の研究では，画像処理技術を使ってキーフレームの抽出と吹き出し配置を行うことで，フィルムコミックの自動生成を実現した．しかし，フィルムコミックの題材となるアニメーション映画では，キャラクターが人間でなかったり，または顔が実写とは異なりデフォルメされているため，彼らの顔検出や唇の検出による話者特定が適用できない場合が多い．一方で，多量のテンプレートから適切なものを選んでコミックにレイアウトする手法[6)]が提案されており，コミックに使用したい画像と画像内の重要箇所のおおよその大きさを何らかの方法

で与えることができれば，この手法を利用することが可能である．

これらに対して，国広ら[7]はフレームの選択には画像処理技術を，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手法を提案した．フレーム選択では，色の変化を検出することで映像を似たシーンに分割し，その中央に位置する画像をシーンの代表フレームとして選択した．そして，鑑賞者が注視した箇所をフレーム内の重要箇所とみなし，それがトリミングと吹き出しの配置により切り取られたり，隠されたりしないようにした．しかし，国広ら[7]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) 吹き出し配置が重要箇所にかかってしまう，といった問題があった．本論文では，国広ら[7]の手法におけるこれらの問題を提起し，これを解決する手法を提案する．

2. 関 連 研 究

初期のフィルムコミックの自動生成は，Hwang ら[1),2)]によって行われた．彼らは，コミックらしい吹き出しの自動配置アルゴリズムを提案している．しかし，話者とセリフの対応付けがなされていることが前提条件であり，フィルムコミックの自動生成におけるメイン処理である，フレーム選出は手動で行う必要がある．Preußら[3)]は映画の脚本内容の書かれたシナリオ情報から，映画をコミックに自動変換する手法を提案したが，一般視聴者にとって入手が困難な映画のシナリオ情報を必要とするため，用途が限られる．近年では，Hong ら[4)]が顔検出や唇検出，モーション解析技術を用いて話者とセリフを対応付け，キーフレームとなるシーンを抽出することで，自動的に映画からコミックを生成する手法を提案した．しかし，顔検出や唇検出では実際の人間の顔を検出対象として想定しているため，この技術を人間以外の生き物をモデルとした，または人間を大きくデフォルメしたアニメーションのキャラクターに適応させることは困難である．近年の研究として Cao ら[6)]は，コミックに載せたい画像とその画像の望ましい大きさを入力として，多量に用意されたコマのテンプレートから適切なものを選択し，コミックにレイアウトする手法を提案した．しかし，彼らの提案手法では，重要なフレームの選出や，フレーム内の重要箇所の選出を手動で行わなくてはならない．また，Toyoura ら[5)]は，映像における代表的なカメラワークを検出し，それをコミック的な表現に変換する方法を提案した．

一方，動画要約[8)]の分野も，フィルムコミックの生成に大きく関連する．動画要約もフィルムコミックの生成も，主に二つの共通した問題が挙げられる．一つ目は，映像からどのようにして最適な画像を抽出するかという問題である．二つ目は，映像の持つ内容をどのように効果的に描写するかという問題である．動画要約の技術はここ 10 年程で飛躍的に進歩し，この問題解決を図るためにコンピュータビジョン[9),10)]やテキストマイニング[11)]，fMRI[14)]にまで研究の幅を広げている．コンピュータビジョンを利用する技術には Saliency Map[13)]や顔認識，Speaking Lip 検出[4)]を用いる方法が提案されている．Saliency Map とは，静止画や動画などにおいて，人の視覚的注意を引きやすい領域を画像処理により求める手法であり，色やテクスチャ方向などの画像特徴を組み合わせることで求められる．Saliency Map[13)]を用いて重要フレームを抽出する手法に関しては，映像視聴時の鑑賞者の興味はカメラワークによる演出，セリフやストーリーの展開にも大きく影響されるため，色やテクスチャ方向などの低次の画像特徴の中心周辺差分により算出される Saliency Map では推定できない場合が多いと言える．またフィルムコミックに登場するキャラクターは動物であったり大きくデフォルメされている場合が多いため，従来の顔や Speaking Lip 検出器がほとんど適用できない．テキスト解析による方法では逆に映像による演出の効果が考慮されていない．fMRI は脳の活動に関連した脳血流の変化を計測する装置であり，これを用いることで鑑賞者に映像を見せながら，脳が内容に応じて反応した箇所を抽出できる．fMRI は鑑賞者の興味をより正確に捉えられるが，装置が高価のため，現在のところその応用が限られる．Peng ら[19)]は，視線情報と顔の表情を検出することで，ホームビデオにおける重要フレームの選出方法を提案した．表現方法についても，コミック調に編集することで，要約結果を効果的に表現する方法が提案されている[20)]．しかし，動画要約は主に動画内容の俯瞰やそれを索引として必要な情報を簡単に検索できるようにすることを目的としている．一方で，フィルムコミックは映画に代わる媒体として利用され，読者が元の映画がなくても容易にストーリーを把握できなくてはならない．したがって，動画要約で提案されている多くの技術をフィルムコミックの作成に直接用いることはできない．例えば，Peng ら[19)]はキーフレームの選出において，長い注視状態が続いている箇所を重要フレームと定義したが，ストーリーの展開を把握するのには，内容が切り替わった箇所をキーフレームとした方がよい．国広ら[7]による，フレーム選択の手法は，シーンの切り替わりを画像内の色のヒストグラムの大きな変化により検知するものであった．しかし，この手法では，ヒストグラムに変化が少ないながらも，登場人物のわずかな表情の変化などストーリーを理解するうえで重要な情報を含むシーンを検知できない．また，国広ら[7]は，鑑賞者の視線は興味のある対象に向けられるため，重要領域や発話主体に関する情報を含んでいると考え，トリミングや吹き出し配置に対して視線を利用した．しかし，彼らの手法では視線のみを使っているため，視線のノイズに生成結果が左右されるのみでなく，会話の相手や重要なコンテキスト情報がトリミングで切り取られたり，吹き出しで覆い隠されたりすることがあった．以上で述べた関連研究を，フィルムコ

ミックを作成するために必要な技術手法とその醸成条件という観点から分類し，表 1 にまとめた．また，関連研究における問題点を，以下に列挙する．

- 国広ら[7]による，フレーム選択の手法では，ヒストグラムに変化が少ないながらも，ストーリーを理解するうえで重要となるシーンが検出できない．
- 国広ら[7]による重要領域の検出法では，視線のノイズに影響されやすく，正しくトリミングやセリフ配置が成されないことがある．
- Hong ら[4]の顔検出や唇検出は，実際の人間を検出対象としているため，フィルムコミックの主な対象である，アニメーションや CG のキャラクターには適応できないので，自動的に発話対象を検出するのは困難である．

3. 提案手法の概要

2 章で列挙した既存研究に対する問題を解決するための，提案手法でのアプローチを以下に述べる．まず，国広ら[7]のフレーム選択の手法では，画像内のヒストグラムの大きな変化によって，重要なフレームを選出していた．しかし，この手法では，ヒストグラムに変化が少ないながらも，登場人物のわずかな表情の変化など，ストーリーを理解するうえでの重要情報が検知できない．そこで本研究では，視線情報は鑑賞者のストーリー理解の過程を推定できる，より高次の特徴を含んでいると考え，フレーム選択の際にも，鑑賞者の視線情報を利用することとした．とりわけ，短時間における，視線の大きな移動であるサッカード (saccade) に注目し，サッカードが生じるタイミングは鑑賞者の興味がシフトしたタイミングであると考えた．一方で，視線情報には意味の無い動きなど，多くのノイズが含まれる．これらを，鑑賞者が意識的に対象を注目したケースと区別する必要がある．この問題を解決するために，本研究では視線情報と画像特徴を併用する手法として iMap(informative Map) を提案する．iMap は，鑑賞者の視線情報と画像における SURF(Speeded Up Robust Feature)[12]を求めて作成され，画像内の情報量の分布を表す．SURF[12]は，画像の拡大・縮小や回転に不変な局所特徴量を得る手法として用いられる．映像作品では，演出によってズーム等が用いられるため，演出効果に頑健な局所特徴量として SURF[12]を使用する．通常，画像内の意味のある箇所には画像特徴が含まれるため，視線が向けられた箇所の周囲に十分な画像特徴があるかを調べることで，視線に含まれるノイズを除去できる．

次に，国広ら[7]による重要領域の検出法では，視線のノイズに影響されやすく，正しくトリミングやセリフ配置が成されないという問題点を挙げる．この問題に対しても，iMap を用いることで解決を試みる．iMap には視線情報だけでなく画像特徴も含まれているため，たとえ視線が意味の無い箇所を向いていたとしても画像自身が持つ重要箇所を切り取ったり，または吹き出しで隠してしまったりするようなケースを避けられる．

最後に，Hong ら[4]の顔検出や唇検出を用いたコミック作成法は，実際の人間を検出対象としているため，フィルムコミックの主な対象である，アニメーションや CG のキャラクターへの適応が難しいという問題を挙げる．提案手法ではこの点を考慮し，特定の物体だけに依存するような特徴量を使用しないため，どのような映画にも対応できる．そして，注視情報を画像特徴と合わせて使用し，映像の中の情報量の分布図を求めることで，最も情報の少ない箇所に吹き出しを配置することを可能にしている．また，鑑賞者の視線は通常話者に向けられるため，注視情報を利用することである程度話者の位置を推定できる．

本研究で提案する手法の概要を図 1 に示す．まず，フィルムコミックの作成に必要な情報として，動画情報，字幕情報，動画視聴時の鑑賞者の視線情報を用意する．字幕情報には，セリフの内容とそのセリフの開始時刻・終了時刻が含まれている．視線情報は視線追跡装置を用いて取得する．次に，動画情報を動画要約技術によって解析し，似たフレームを持つショットに分ける．さらに，iMap を利用して，このショットを細分化する．その後，ショットごとにカメラワークの有無を判定する．こうして，ショットごとの開始時刻と字幕情報とを統合し，各ショットの開始・終了時刻と字幕情報を持つ脚本データを作成する．続いて，Toyoura ら[5]の手法を利用して，前後の演出効果などからコミックのコマ割り設計を行う．Toyoura ら[5]の手法では，映像における演出効果に合わせて用意されたコマのテンプレートレイアウトを使用する．そして，脚本データから使用する画像を選択し，iMap を用いてコマにサイズを合わせるためのトリミングを行う．さらに，脚本データの演出情報を用いることで，マンガ独特の演出効果を加える．最後に，iMap を用いて吹き出しの位置を決定し，フィルムコミックとして表示する．

4. 実 現 方 法

本研究では，国広ら[7]の研究に対して改善を行うことで，(1) ストーリーの内容を理解するための重要なフレームの選出，(2) 元の映画の構図を保存しつつ重要箇所を切り取らないトリミング，(3) 読みやすく重要箇所を隠さない吹き出し配置を実現する．以降の節では，国広ら[7]の手法に対する問題点を述べ，これを解決する手法を提案する．

4.1 iMap の作成

フィルムコミックを作成するためには，映像の内容を理解し，その内容をよく表すように画像をトリミングする必要がある．この問題は，画像上のそれぞれの領域がどれほどの情報を持っているかを計算することで解決できる．我々は，鑑賞者の視線位置と画像特徴から画像上の各点での情報の量を記述する iMap を生成し，これを実現する．

視線追跡装置から視線位置を獲得できる．多くの鑑賞者か

表 1　フィルムコミック作成に必要な技術手法と関連要素技術

Table 1　Related works summarized as the 3 tasks required for film comic generation(column) and the elemental technologies(row)

関連要素技術	フレーム選出	レイアウト	セリフ配置
ユーザ入力	[1],[2],[3],[6]	[3],[6]	[1],[2],[3],[6]
顔・唇検出	[4],[19]	[4]	[4]
モーション解析	[8],[9],[10],[20]	[20]	[20]
Saliency Map	[13]		
視線	[19]	[7]	[7]
fMRI	[14]		
テキストマイニング	[3],[11]		[3]
カメラワーク検出		[5]	
iMap	[本稿]	[本稿]	[本稿]

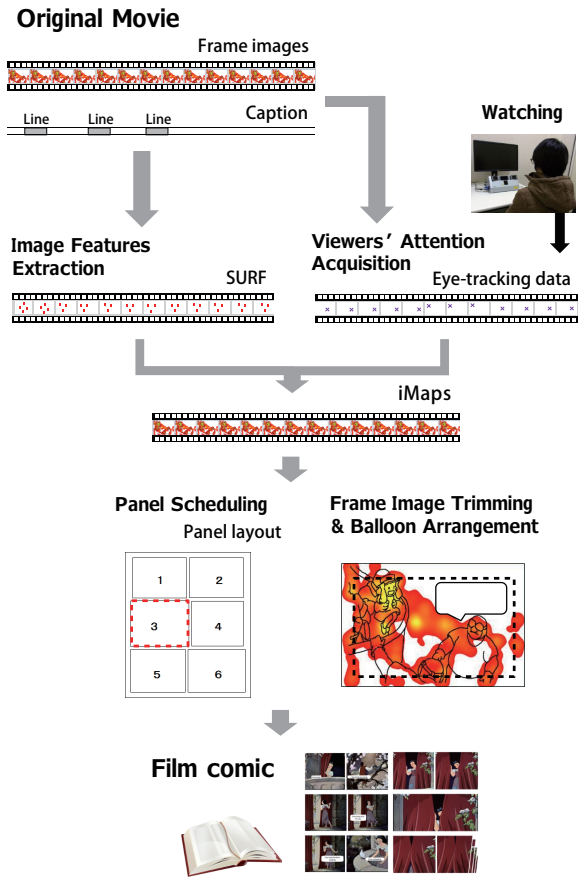


図 1　フィルムコミック作成手法概要

Fig. 1　The framework of proposed film comic generation method.

ら長く注視される位置は関心のある情報を含む可能性が高い．t フレームに対して得られる n 番目の鑑賞者の視線位置から，t フレームにおける注視マップ M_a^t を以下のように定義する．ここで注視マップ M_a^t は図 2 中の Attention map の事を指す．

$$M_a^t = \sum_{i=1}^{n} f_{i_{EYE}}^t(x,y) \otimes G(0,\sigma_a^2) \tag{1}$$

$$f_{i_{EYE}}^t(x,y) = \begin{cases} 1 & if\ (x,y)\ is\ an\ eye\ position \\ 0 & if\ (x,y)\ is\ not\ an\ eye\ position \end{cases}$$

ただし，$G(0,\sigma_a^2)$ は平均 0，分散 σ_a のガウスカーネルを示し，$\otimes$ は畳み込み演算を行うことを示す．σ_a は任意の定数とする．

国広ら[7]の手法では，用いる視線情報は被験者一人分であるため，被験者の視線位置が必ずしも発話者を捉え，重要箇所を示すわけではなかった．このことは，トリミングや吹き出し配置において，問題となることが多かった．そこで我々は，複数の鑑賞者の視線位置を注視マップという形で統合することとした．

また，画像内の情報をもつ領域からは画像特徴点が検出されると期待できる．本研究では画像特徴点として SURF[12]を用い，画像上のそれぞれの点でどれほどの情報の量が見込まれるかを計算する．t フレームにおける画像特徴マップ M_f^t を以下の式により定義した．ここで画像特徴マップ M_t^t は図 2 中の Feature map の事を指す．

$$M_f^t = f_{SURF}^t(x,y) \otimes G(0,\sigma_f^2) \tag{2}$$

$$f_{SURF}^t(x,y) = \begin{cases} 1 & if\ (x,y)\ is\ SURF \\ 0 & if\ (x,y)\ is\ not\ SURF \end{cases}$$

$G(0,\sigma_f^2)$ は平均 0，分散 σ_f のガウスカーネルを示す．σ_f は任意の定数とする．

最後に図 2 のように，注視マップ M_a^t と画像特徴マップ M_f^t の両方を組み合わせ，画像における情報の量を表す iMap M^t を作成する．

$$M^t = M_f^t * M_a^t \tag{3}$$

ただし，$*$ は画素ごとの掛け算を表す．

iMap の作成に用いる視線情報は，フレーム選択時とトリミング，吹き出し配置時において異なる．フレーム選択時はそれぞれのフレームについて iMap を作成し，iMap に十分

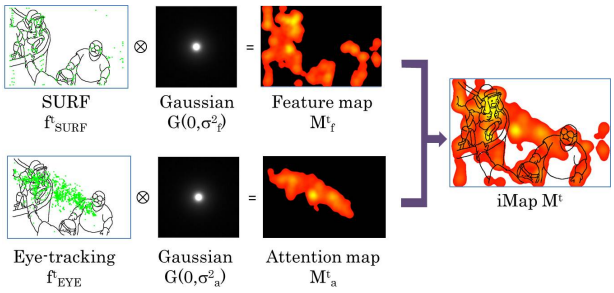


図 2 視線位置と画像特徴点からの iMap の作成
Fig. 2 Creation of iMaps by aggregating eye-tracking points and image feature points.

な量の変化が生じた瞬間のフレームを選出する．一方で，トリミング，吹き出し配置時は，選ばれたフレームに対してその前後の iMap を足し合わせる．これにより，複数のキャラクター間で会話やインタラクションがある場合，話者や動作する側だけでなく，聞き手や動作を受ける側も考慮した興味領域の抽出が可能となる．

4.2 フレーム選択

国広ら[7]は，フレーム選択を画像処理技術のみで行っている．カット検出を色のヒストグラムの大きな変化で検出し，ショットの中央フレームを代表フレームとしてコマに使用した．しかし，この手法には，(1) 内容の変化や動作主体の変更といった文脈依存の情報を検知できない，(2) 色のヒストグラムの変化を検知する際の閾値設定が困難である，といった二つの問題がある．例えば，背景や人物が同じであれば，人物の位置や動作に大きな変更があっても，色のヒストグラムは変化しないため，この方法ではストーリーの展開において重要なフレームを選出できない可能性がある．また，色のヒストグラムの変化を検知する際の最適閾値は，シーンごとに異なるはずであるため，自動的に定めることが困難である．本研究では，視聴者の視線は通常キャラクターの動きや内容の変化に追随しているという考察に基づき，視線を利用してフレームの検出を行う．

既存の動画要約の研究において，鑑賞者が長く注視していたフレームを重要フレームとして選出する方法が提案されている[19]．データの圧縮や内容の要約が重要であるビデオ要約においてはこのような方法が有効であるが，フィルムコミックはストーリーの展開の表現が重要であるため，提案手法では画像処理によって得られる各ショットに対して，ショット内の前後のフレームの iMap の差分を計算し，差分が一定の閾値を超えたところでショットを細分割する．最後に各サブショットの先頭のフレームを重要フレームとして選出する．提案手法で使用する動画の長さに制約はなく，また抽出されるコマの枚数は iMap の変化を検知する閾値に依存する．ユーザのニーズに応じて，パラメータを調整することで，入力動画から出力されるコマ数を調節することが可能となる．

4.3 トリミング

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．フレーム画像には映画製作者の意図した場面の見せ方を反映した構図がある．構図を保ったままトリミングを行うには，編集者のセンスや長い経験が必要であり，これを自動的に行うことでフィルムコミック作成のコストを減らすことができる．

国広ら[7]はフレーム画像中の重要箇所を切り取らないように，鑑賞者の視線情報を利用してトリミングを行った．彼らの手法では，まずショット内の視線の位置をクラスタリングして，最大クラスタの重心を求める．そして，この重心が切り取った後にコマの中央にくるように，画像をトリミングする．この方法において，フレーム画像におけるこの重心の位置が画像の中央でない場合は，トリミングによりフレーム画像の構図が壊されてしまう可能性がある．この問題を解決するために，提案手法では重要箇所が切り取られていないか，かつフレーム画像における重要箇所の相対位置が保存されているかを評価するエネルギー関数を，iMap を用いて定義し，最適化問題として解くことで最適なトリミングを実現する方法を提案する．

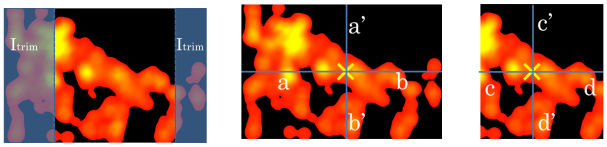
重要箇所が切り取られないことを評価するために，トリミングによって切り取られてしまうコスト E_{elim} を式 (4) により計算する．

$$E_{elim}=\sum_{(i,j)\in I_{trim}} a_{ij} \tag{4}$$

ここで，a_{ij} は切り取られる画素での iMap の値である．I_{trim} は図 3(a) に示すように，切り取られる領域を示している．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 (5) のように構図のコスト E_{str} を定義する．ここで，a，b は注目画素のトリミング前の横方向の比率を示し，c，d は注目画素のトリミング後の横方向の比率を示す．また，a’，b’ は注目画素のトリミング前の縦方向の比率を示し，c’，d’ は注目画素のトリミング後の縦方向の比率を示している．図 3(b)，(c) 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．

$$E_{str}=\sum_i\sum_j a_{ij}\times\left|\left(\frac{a}{b}-\frac{c}{d}\right)+\left(\frac{a'}{b'}-\frac{c'}{d'}\right)\right| \tag{5}$$

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．提案手法では，探索するトリミングのサイズを動的に変化させながら最適なトリミングを求めていく．すなわち，コマのアスペクト比を保ったまま，切り取る領域を大きくしながら探索を行う．これにより，最適な



(a) (b) (c)

図 3 トリミング実現のための領域コストの計算．(a)E_{trim} のコスト，(b) フレーム画像における画素 a の相対位置，(c) トリミング後の画像における画素 a の相対位置．

Fig. 3 Calculating the cost for trimming an image. (a) Cost E_{trim}, (b) the position of pixel a in a frame image, and (c) the position of pixel a in the trimmed image.

構図を保ったまま，重要でない領域をできるだけ多くカットすることができる．その結果，コマに画像をはめ込むときに拡大され，重要箇所のみを拡大して表示する効果が得られる．

4.4 吹き出し配置

セリフのあるコマでは，発話者を推定しながら，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．この問題に対し提案手法では，鑑賞者の視線情報と画像の局所特徴から発話対象を推定し，発話対象に近く，また重要箇所を隠さない位置に，読み易いセリフ形状で，自動的にセリフを配置する手法を実現する．国広ら[7]は，視線情報が多く集まる上位 2 つの注視領域から線を引きコマを 9 つの領域に別け，その領域中で最大の面積を持つ領域に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．例えば，図 7(a) に示す既存手法の例では，吹き出しが発話者本人を隠してしまっている．これは，視線の向く先は発話主体である可能性が大きいが，必ずしも発話者であるとは限らないことによる問題である．

これに対し，提案手法ではもっとも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領域 $I_{balloon}$ に含まれる画素の iMap 値の合計を式 (6) 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 (7) のように吹き出しの重心 (x_g, y_g) から，iMap においてもっとも重要度の高い位置 (x_a, y_a) までの距離 E_{dist} もコストに加える．さらに，吹き出しの形状が極端に細長くならないように，コストに式 (8) で示す吹き出しのアスペクト比 E_{shape} の項も含める．式 (8) 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．吹き出しの横の長さと縦の長さは，セリフ情報を適切な位置で改行しながら反復施行を行うことで求められる．

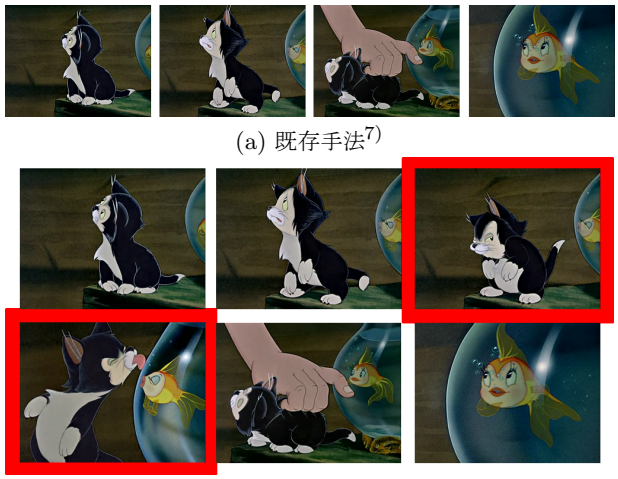
$$E_{hide} = \sum_{(i,j)\in I_{balloon}} a_{ij} \tag{6}$$

(a) 既存手法[7]

(b) 提案手法

図 4 フレーム選択結果例 1

Fig. 4 An example of selected frames by our proposed method. (Case 1)



(a) 既存手法[7]

(b) 提案手法

図 5 フレーム選択結果例 2

Fig. 5 An example of selected frames by our proposed method. (Case 2)

$$E_{dist} = \sqrt{(x_a - x_g)^2 + (y_a - y_g)^2} \tag{7}$$

$$E_{shape} = \left| \frac{L_{width}}{L_{height}} \right|^2 \tag{8}$$

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，読み易い吹き出しを配置できる．吹き出しの位置が決まったら，iMap において高い値を持つ領域を求め，その方向へと吹き出しの尾を配置する．iMap における高い値を持つ領域が一箇所に特定できない場合には吹き出しの尾を配置しない．

5. 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 代の男性 3 名であり，Disney 製作の「白雪姫」,「ピノキオ」の冒頭 20 分を見てもらった．この 3 名の視線情報を用いてフィルムコミックを作成した．

5.1 フレーム選択

図 4，5 に，提案手法と国広らの手法[7]による結果を示す．国広らの手法を用いた結果では，画素に大きな変化が現れない限りカットと判断されないため，ストーリーを理解する上での重要情報が抜け落ちている．図内の赤枠で囲ってある画像が，提案手法で新たに検出できた画像を示す．

図 4 の例において，このシーンでは狩人が白雪姫を殺そうとしているが，良心の呵責に苦しみ，結果として白雪姫を殺すことはできない．(a) の既存手法の結果では，狩人がためらうシーンがない．一方で，(b) の提案手法の結果では狩人のためらう表情がよく抽出されている．このコマがあることによってセリフには現れない，重要な情報が反映される．

図 5 の例において，このシーンでは猫のフィガロと金魚のクレオが写っている．もう夜も遅いので，寝ようとゼペットじいさんが言った直後のシーンである．ゼペットじいさんは，フィガロにクレオに対しておやすみのキスをするように言う．しかし，フィガロは照れているのか，なんだか不機嫌な様子で，クレオの入っている金魚鉢をペロリと舐めて，ゼペットじいさんに連れられていく．

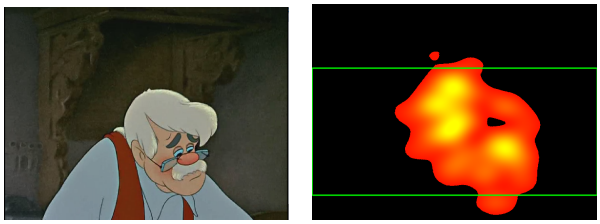
(a) の既存手法では，フィガロがクレオにキスをするシーンが抜けてしまっている．一方で，(b) の提案手法では，フィガロが一瞬，体を固くするような仕草や，クレオにキスをするシーンがしっかりと取れている．このシーンでは， キスをするという重要なジェスチャーが抽出できないと，前後のシーンの意味が繋がらなくなってしまう．

このように，ストーリーの展開を表す重要なフレームが頻繁に欠けると，読者はコマとコマの間を推測しなければならず，内容の把握が困難となる．特に，人のボディランゲージやジェスチャーなどの非言語表現では，色のヒストグラムに大きな変化をもたらさないため，国広ら[7]の手法では表現しきれない．それに対して提案手法はそれに追随して移動する視線を検知することにより，効果的に表現することができる．こうした非言語表現は，ストーリーを理解するうえで重要な手がかりになるため，これらの検出の実現によって映像作品の魅力をコミックにより反映させられる．

実際のフィルムコミックの製作においては編集者が，膨大なフレームの中から使用するフレーム画像を選ぶ．同様のシーンが多すぎると単調になり，逆に少なすぎると，ストーリーが表現しきれないため，生成されるフィルムコミックの質は編集者の経験，スキルと個人の好みに大きく依存する．それに対して，提案手法は複数人の視線データを利用することにより，より客観的な視点からみて冗長ではなくかつストーリーを理解する上で必要な情報を持ったフレーム画像を選ぶことができる．また，提案手法によって検出されたフレームが妥当であるかを評価実験により確かめた．被験者は，映画を鑑賞した被験者を含まない大学生８名（男性４名，女性４名）であり，まず使用した動画とは別作品の，市販されているフィルムコミックを読んでもらった．そして，国広らの手法で検出されるフレーム画像に対して，提案手法により増加して検出された各フレーム画像が，本物のフィルムコミックのコマの選出方法と比べて必要なフレームの増加であったかどうかを，①必要，②あってもなくてもよい，③必要でない，の三段階尺度によって増加した全てのフレーム画像に対して評価してもらった．帰無仮説を”①と回答する場合とそれ以外を回答する場合の確率は等しい”とし，二項検定によって分布の偏りを調べた．有意水準 5%で有意差が見られたフレーム枚数を数えたところ,「白雪姫」では 59 枚の増加フレーム画像に対して，77.9%に当たる 46 枚であった．また,「ピノキオ」では，80 枚の増加フレーム画像に対して，86.2%に当たる 69 枚であった．次に，帰無仮説を”①または②と回答する場合とそれ以外を回答する場合の確率は等しい”とし，二項検定によって分布の偏りを調べた．有意水準 5%で有意差が見られたフレーム枚数を数えたところ,「白雪姫」では 59 枚の増加フレーム画像に対して，94.9%に当たる 56 枚であった．また,「ピノキオ」では，80 枚の増加フレーム画像に対して，97.5%に当たる 78 枚であった．

5.2 トリミング

国広ら[7]の手法で用いられているコマの形状とサイズはテンプレートによって決定する．また，今回我々が使用した Toyoura ら[5]の手法も，コマの形状とサイズは映像における演出効果に合わせて用意されたテンプレートレイアウトを使用する．矩形のコマが使用されるため，フレームのアスペクト比とコマのアスペクト比が同じである場合は，フレームをコマに合わせてリサイズするのみで処理が済む．トリミングが必要となるのは，アスペクト比が異なる場合であり，アスペクト比が大きく異なるほど，国広ら[7]の手法に比べて，本手法の効果がより顕著となる．国広らの手法[7]と提案手法によるトリミング結果を図 6 に示す．図 6 の例では，元のフレーム画像とコミックのコマのアスペクト比が大きく異なるため，トリミングにおける影響が大きくなる．オリジナル画像 (a) に対して，生成された iMap が (b) であり，(b) 内の緑色の枠は，トリミング位置を示している．(c) の既存手法ではゼペット爺さんの顔が完全に見切れてしまっている．これは既存手法のように一人の視線情報のみに頼ると，重要箇所が正しく推定されず，視線がもっとも集中する位置が中央にくるようにトリミングしてしまうと不自然なトリミング結果となることを示している．一方で，(d) の提案手法では (b) の iMap に示されているように，画像特徴と複数人の視線情報をシーンの前後を考慮して用いているため，ゼペット爺さんの顔が見切れることはない．しかし，トリミングが必要となるのは，あくまでも映画とコミックのコマのアスペクト比が大きく異なる場合である．今回の実験では，映画のサイズとコミックのコマのサイズのアスペクト比が大きく異なることが少なかった．

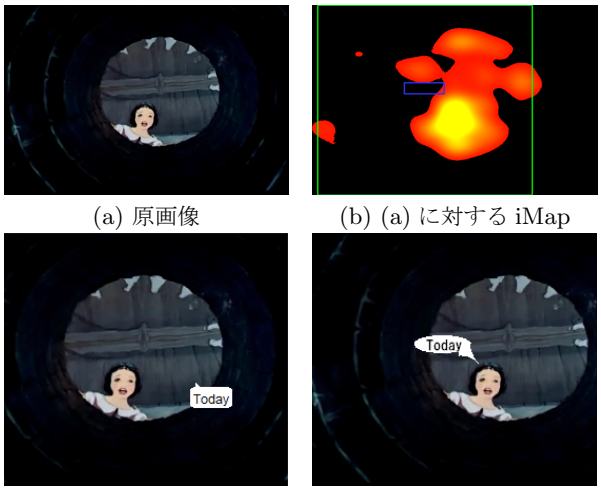


(a) 元のフレーム画像　(b) (a) に対する iMap

(c) 既存手法[7]　(d) 提案手法

図 6　トリミング結果例

Fig. 6　An example of trimming an image by our proposed method.



(a) 原画像　(b) (a) に対する iMap

(c) 既存手法[7]　(d) 提案手法

図 7　吹き出し配置結果例 1

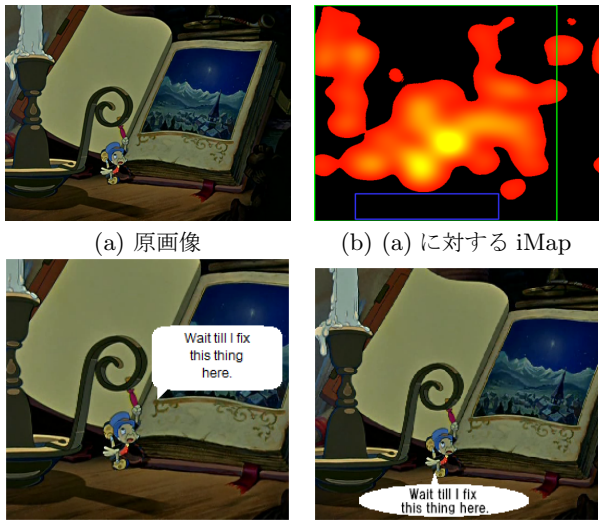
Fig. 7　An example of balloon arrangement by our proposed method. (Case 1)

5.3　吹き出し配置

図 7，8 に国広ら[7]による既存手法と提案手法で作成した吹き出し配置の例を示す．

図 7 において，オリジナル画像 (a) に対して生成された iMap が (b) であり，(b) 内の緑色の枠はトリミング結果を，青色の枠が吹き出し位置を示している．(c) の既存手法の結果では，吹き出しは視線が集中した箇所付近に配置され，吹き出しの尾もそこを指している．しかし，実際に視線が向いた先が発話者では無かった例である．これに対して，(d) の提案手法では白雪姫に向けて適切に吹き出しの尾が向けられ，また吹き出し位置も重要箇所を隠さない自然な位置である．これは，(b) の iMap を見てわかるように，複数人の視線から発話者と思われる位置を推定し，また SURF を用いることで周りの特徴のないところの重要度が相対的に低く抑えられたためである．

図 9 の例では，(c) の既存手法の結果は，話者に対しては適切に吹き出しが配置されているが，このシーンにおける重



(a) 原画像　(b) (a) に対する iMap

(c) 既存手法[7]　(d) 提案手法

図 8　吹き出し配置結果例 2

Fig. 8　An example of balloon arrangement by our proposed method. (Case 2)

要箇所を隠してしまう結果となっている．

一方で，(d) の提案手法では，重要箇所を隠すことなく，かつ正しい話者の位置に吹き出しが配置されている．これは，(b) の iMap に示されるように，画像特徴と複数人の視線のある程度長い時間の視線位置から，話者の候補とこのシーンの前後の重要箇所をうまく推定できるためである．このシーンでは，コオロギが絵本を差しながら説明を始めるため，視線情報もコオロギから本へと誘導される．(b) の iMap では，コオロギだけでなく絵本も重要領域となっていることが確認できる．

提案手法と既存手法でそれぞれ作成したフィルムコミックの吹き出し配置におけるエラー率を求めたところ，次のようになった．白雪姫の冒頭 10 分間で吹き出しは 94 個生成された．既存手法では 4 個のエラーが存在し (4.2%)，提案手法ではエラーが存在しなかった．ピノキオの冒頭 20 分間で吹き出しは 243 個生成された．既存手法では 19 個のエラーが存在し (7.8%)，提案手法ではエラーが存在しなかった．ここでは，第三者にセリフが重要箇所を隠してしまっていないか，またセリフの尾の方向が適切な発話者に向いているかを評価してもらい，以上が不自然な吹き出しをエラーとしてその枚数を数えた．

6.　おわりに

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法を提案した．鑑賞者の視線位置と画像特徴から画像上の重要箇所を示す iMap を作成することで，ストーリーの表現に重要なフレームの選択を可能にし，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．フィルムコミックのターゲットとする人の年齢や，ニーズに合わせて使用する被験者を変えることで，

ターゲットとする読者に適応したコミックを作成できることは，本研究の利点の一つであると考える．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の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．さらに，コミックのレイアウトに関してもテンプレートではなく，選ばれたフレーム画像の内容によってコマサイズを決定するなどの発展も考えられる．その場合，コマのアスペクト比が元の映画と大きくことなるケースが増えるため，我々の提案手法の有効性がさらに示されると考える．

参 考 文 献

1) B.K. Chun, D.S. Ryu, W.I. Hwang, and H.G. Cho, “An automated procedure for word balloon placement in cinema comics,” International Symposium on Visual Computing (ISVC), vol.2, pp.576–585, 2006.
2) W.I. Hwang, P.J. Lee, B.K. Chun, D.S. Ryu, and H.G. Cho, “Cinema comics: Cartoon generation from video stream,” International Conference on Computer Graphics Theory and Applications, pp.299–304, 2006.
3) J. Preuß and J. Loviscach, “From movie to comics, informed by the screenplay,” ACM SIGGRAPH (Poster), 2007.
4) R. Hong, X.T. Yuan, M. Xu, M. Wang, S. Yan, and T.S. Chua, “Movie2comics: a feast of multimedia artwork,” Proceedings of the International Conference on Multimedia, pp.611–614, 2010.
5) M. Toyoura, M. Kunihiro, and X. Mao, “Film comic reflecting camera-works,” International Conference on MultiMedia Modeling, pp.406–417, 2012.
6) Y. Cao, A.B. Chan, and R. Lau, “Automatic stylistic manga layout,” ACM Transactions on Graphics (Proc. of SIGGRAPH Asia 2012), vol.31, 2012.
7) 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.
8) A.G. Money and H. Agius, “Video summarisation: A conceptual framework and survey of the state of the art,” Journal of Visual Communication and Image Representation, vol.19, no.2, pp.121–143, 2008.
9) P.P. Agouris and P. Doucette, “Summarizing video datasets in the spatiotemporal domain,” International Workshop on Database and Expert Systems, Applications, pp.906–912, 2000.
10) S.V. Porter, “Video segmentation and indexing using motion estimation,” PhD thesis, University of Bristol, 2004.
11) B. Chen, J. Wang, and J. Wang, “A novel video summarization based on mining the story-structure and semantic relations among concept entities,” IEEE Transactions on Multimedia, vol.11, no.2, pp.295–312, 2009.
12) H. Bay, A. Ess, T. Tuytelaars, and L.V. Gool, “Speeded-up robust features (surf),” Computer Vision and Image Understanding, vol.110, no.3, 2008.
13) S. Marat, M. Guironnet, and D. Pellerin, “Video summarization using a visual attention model,” European Signal Processing Conference (EUSIPCO), pp.1784–1788, 2007.
14) X. Hu, F. Deng, K. Li, T. Zhang, H. Chen, X. Jiang, J. Lv, D. Zhu, C. Faraco, D. Zhang, et al., “Bridging low-level features and high-level semantics via fmri brain imaging for video classification,” Proceedings of the international conference on MultimediaACM, pp.451–460 2010.
15) Y.F. Ma, X.S. Hua, L. Lu, and H.J. Zhang, “A generic framework of user attention model and its application in video summarization,” IEEE Trans. on Multimedia, vol.7, no.5, pp.907–919, 2005.
16) K. Li, L. Guo, C. Faraco, D. Zhu, F. Deng, T. Zhang, X. Jiang, D. Zhang, H. Chen, X. Hu, S. Miller, and T. Liuothers, “Human-centered attention models for video summarization,” International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal InteractionACM, p.27 2010.
17) A. Liu and Z. Yang, “Watching, thinking, reacting: A human-centered framework for movie content analysis,” International Journal of Digital Content Technology and its Applications, vol.4, no.5, pp.23–37, 2010.
18) J.Y. You, G.Z. Liu, L. Sun, and H.L. Li, “A multiple visual models based perceptive analysis framework for multilevel video summarization,” IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no.3, pp.335–342, 2007.
19) W.T. Peng, W.J. Huang, W.T. Chu, C.N. Chou, W.Y. Chang, C.H. Chang, and Y.P. Hung, “A user experience model for home video summarization,” International Conference on Multimedia Modeling, pp.484–495, 2009.
20) J.S. Boreczky, A. Girgensohn, G. Golovchinsky, and S. Uchihashi, “An interactive comic book presentation for exploring video,” ACM Conference on Computer-Human Interaction (CHI), pp.185–192, 2000.

(2013 年 2 月 4 日受付)
(2013 年 6 月 17 日再受付)

澤 田 友 哉
2012 年 山梨大学工学部卒業．2012 年 同大学大学院医学工学総合教育部修士課程 入学．現在に至る．映像処理，コンピュータグラフィックスの研究に従事．

豊 浦 正 広
2003 年 京都大学工学部情報学科 卒業．2007 年 日本学術振興会特別研究員 DC2．2008 年 京都大学大学院情報学研究科博士後期課程 修了．2008 年 日本学術振興会特別研究員 PD．カリフォルニア大学サンタバーバラ校訪問研究員．2009 年 山梨大学大学院医学工学総合研究部 助教．拡張現実感，コンピュータビジョンの研究に従事．電子情報通信学会，情報処理学会，IEEE 各会員．博士 (情報学)．

茅 暁 陽 （正会員）
1983 年 中国復旦大学 計算機学科卒業．1987 年 東京大学大学院情報科学研究科修士課程 修了．1990 年 東京大学大学院理学研究科博士課程 修了．1990 年 株式会社クボタコンピュータ応用エンジニア．1994 年 New York 州立大学 Stony Brook 校 客員研究員．1995 年 科学技術振興事業団特別研究員．1996 年 山梨大学工学部 講師．1997 年 同准教授．1997 年 同大学院医学工学総合研究部 准教授．2008 年 同教授． 画像処理，コンピュータグラフィックス，可視化の研究に従事．情報処理学会，画像電子学会，ACM 各会員．理学博士．