

フィルムコミック形式によるアニメ作品の自動映像要約手法

A Video Summarization Technique of Animation Products According to Film Comic Format

岡寄 堅仁 石川 知一 柿本 正憲 西田 友是

Kento OKAZAKI[†] Tomokazu ISHIKAWA^{†1}

[†] 東京工科大学

[‡] 広島修道大学/UEI リサーチ

Masanori KAKIMOTO[†] Tomoyuki NISHITA[§]

[†] Tokyo University of Technology

[‡] Hiroshima Shudo University / UEI Research

1. はじめに

近年、インターネットによる動画配信などにより、消費者が触れることのできる情報は増加した。特にアニメ作品の動画配信の増加が顕著であり、無料有料問わず、多くの動画配信サイトが存在する。しかしながら、時間や場所などの制約がないインターネットであるだけに、非常に多彩な作品が存在し、動画であるアニメ作品の消費には、まとまった時間が必要であり、すべての人が手軽に楽しめるとは限らない。

このような背景から、映像要約に関する研究も盛んに行われ、必要最低限の情報を抜き出す手法として漫画形式の映像要約も注目されている[1-3]。これらの研究は、動画以外の情報を必要とする限定的な条件であり、シーンの重要度からキーフレームを検出する手法ながら、そこに漫画の技法を適用しておらず、漫画形式の映像要約としては不十分である。

そこで、本研究ではフィルムコミックという書籍メディアに着目した。フィルムコミックは、レイアウトと描画を同時に作成する漫画とは異なり、既にアニメ作品として仕上がったものを漫画形式に編集したものである。フィルムコミックの製作は、編集するアニメ作品の長さが1時間を超えるものが多く存在するため、利用フレームの選択やレイアウトは非常に手間のかかる作業である。

本研究では、フィルムコミックの編集作業を短縮するために、入力したアニメ作品から、漫画技法を考慮した映像要約を行うことを目的とする。入力したアニメ作品から、シーンの切り替え地点のフレームを判定し抽出するだけでなく、コマ割りに利用できる見栄えの良いフレームを検出する。また顔検出を用いて、漫画のコマ割りの形状や大きさとのマッチングを行い、フィルムコミックのようなレイアウト画像を出力するまでの処理を、自動で行えるようにすることを目指す。

2. 関連研究

この節では、漫画レイアウトの関連研究と、映像要約手法、アニメ作品に特化した特徴量の計算についての関連研究を説明する。

2.1 漫画風レイアウトの最適化

Caoらは既存の漫画レイアウトを利用したデータ駆動型のレイアウト作成方法を提案した[4]。確率的生成モデルによって新規にそれらしいレイアウトを合成し、連続する静止画の中からユーザーが指定した領域と配置パネルにおいて最適化を図っている。

レイアウトの最適解を提示した後もフレームを自由に変形できるインターフェースを用意し、初心者ユーザーでも容易に漫画風レイアウトを作成できるが、アニメ作品から使用する画像やトリミングはユーザーが指定する。本研究では、見栄えの良いフレームを自動で選択し、フィルムコミックのレイアウトを自動で生成する方法を提案する。

2.2 アニメ作品の映像要約

福里らは、アニメ作品を漫画として適切に要約するためのキーフレームを、画像から得られる特徴量で判定する手法を提案した[5, 6]。動画をショットごとに分割し、ショット単位、フレーム単位の二階層で重要度を算出し、コマとして相応しいキーフレーム検出を行い、原作漫画との一致度合いを基準とした精度評価の尺度を提案している。

しかし、この研究ではコマのサイズや数を事前に用意したラベリングページのみで対応しているために、キーフレームに対して相応しいコマの生成を行うことが難しい。また、キーフレームの検出を非常に正確に行っているが、必ずしもショット単位から1枚のキーフレームを検出することがストーリーを伝えるための映像要約として正しいとは限らないという問題もある。

2.3 アニメ画像における特徴量抽出

高山らはアニメキャラクターの形状や色の特徴に着目し、キャラクターの顔検出する手法を提案した[7]。顔検出には、肌色検出、エッジ検出、領域分割、輪郭線の二次曲線近似、目の対称性を利用している。肌色がHSV表色系で色相6~38度に集中することに着目し、色相0~40度の範囲を肌色として検出する。

増田らは、画像類似度に基づく類似キャラクターの検索手法について検討し、注目したいキャラクターが存在する動画内のシーンを検索、推奨する手法を提案した[8]。この研究では、ショット分割から、顔検出、特徴量計算、キャラクターの登場シーンの検索を行っている。ショット検出では、隣接フレームの輝度値のヒストグラムを用いた計算を行い、閾値を超えた時にショットの境界を判定する。

3. フィルムコミックの特徴

本研究で着目したフィルムコミックの特徴を説明する。フィルムコミックは、レイアウトと作画を同時に行なう漫画と異なり、作画が完了しているアニメ作品内の静止画を利用し、フレームの配置を行い、効果

¹ ishiawatm@stf.teu.ac.jp

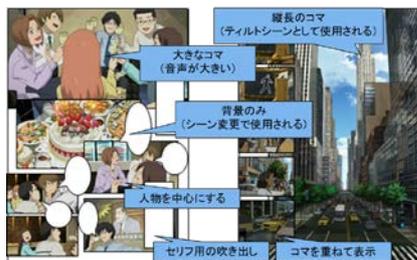


図 1: フィルムコミックの特徴

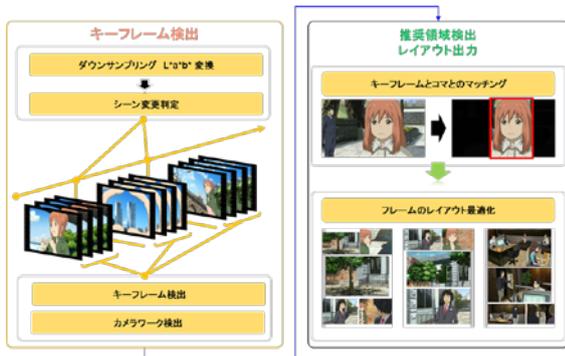


図 2: 提案法の流れ

©プロダクション I.G. 東のエデン劇場版 I The King of Eden. アニミック/フジテレビ. 2010.

音を擬音として、セリフを吹き出しとして加えたもので、アニメ作品を漫画形式で楽しむことができる書籍メディアである。

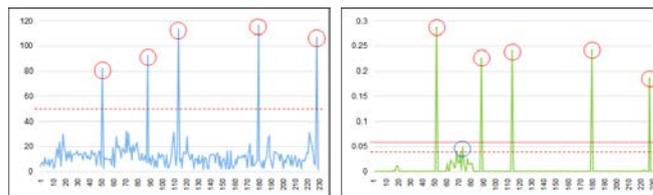
図 1 にフィルムコミックのレイアウト例を示す。以下のような特徴が確認できる。

- ・ 人物が映るシーンでは人物の領域を切り抜いてコマとして用いる傾向にある
- ・ キャラクターが動く様子は、複数コマを用いて表現する
- ・ 背景がメインのシーンでは大きいコマとして描写されることが多い
- ・ アニメ作品内でカメラを上下に動かすことで縦長の大きなコマを写す場合は縦長のコマがそのまま配置される
- ・ 音量の大きいシーンは大きなコマが割り当てられる傾向にある
- ・ 同時系列のシーンではコマを重ねて表現することがある

本研究では、動作に対する複数枚のキーフレーム検出や人物の領域や背景によるコマの切り抜き、タイトルによる縦長のコマの表現、音声を考慮した映像要約を行う。本手法では同時系列の判定と、コマを重ねて描写する技法、また吹き出しや擬音記述の挿入は行わない。

4. 提案法

提案法の処理は「キーフレーム検出」、「推奨領域検出・レイアウト出力」の 2 つに大別される。ここで推奨領域とはコマの中で必ず表示してほしいキャラクターの顔などの重要な領域のことである。提案法の処理の流れを図 2 に示す。キーフレーム検出では、見栄えの良いシーンを選ぶための処理を行う。推奨領域検出では、漫画内の枠であるコマの大きさや形をキーフレーム画像とマッチングするために必要な推奨領域の検出を行う。レイアウト処理では、キーフレーム検出、



(a) ヒストグラム距離判定 (b) ユークリッド距離判定

図 3: シーン変更判定における値の変化。それぞれ横軸はフレーム番号 f で、赤色の丸はシーンが変わったと判定されたフレームを示す。縦軸は、(a)では式(3)で計算されたヒストグラム間の距離 $D(H_f, H_{f-1})$ 、(b)では式(2)で計算される C_f の値を示す。赤い線はそれぞれの値に対応する閾値を表す。

推奨領域検出によって得られた結果から、コマとのマッチング処理を行い、各コマの領域を最小二乗法で最適化し、漫画形式でレイアウト画像を出力するまでの処理を行う。この節ではそれぞれの処理についての詳細を説明する。

4.1 シーン変更判定

アニメ作品に登場するキャラクターなどはそれぞれ個性的な髪の色、服の色、肌の色を持つことが多い。キャラクターに限らずともロボットや乗り物であってもそれぞれ個性的な色を持ち、それらの色はそれぞれを区別するためのアイコンとして視覚的に機能する。そのアイコンとしての色を比較することに着目した。1 フレームをいくつかのブロックに分割し、ブロック内で平均化した色の差で比較するものとした。ブロックは $MN/(mn)$ ブロックで構成される。ここで M, N はそれぞれ入力動画の横のピクセル数、縦のピクセル数、 m, n はそれぞれブロックの横ピクセル数、縦ピクセル数を示す。

次に、ブロック内の各要素の RGB 情報を $L^*a^*b^*$ 表色系に変換する。隣接する 2 フレームの色差 Δ を求め、この色差を用いてシーンの切り替えを判定する。ブロック配列要素 i について、 f 番目フレームと $f-1$ 番目のフレームから得られる評価値を以下の式(1)によって決める。

$$v_{f,i} = \begin{cases} 0 & (\Delta \leq \Delta_{min}) \\ \frac{\Delta - \Delta_{min}}{\Delta_{max} - \Delta_{min}} & (\Delta_{min} \leq \Delta \leq \Delta_{max}) \\ 1 & (\Delta_{max} \leq \Delta) \end{cases} \quad (1)$$

ここで、 Δ_{min} は $v_{f,i}$ に下限の色差、 Δ_{max} は上限の色差を示す。色差 Δ をブロック配列要素 i ごとに計算し、 $v_{f,i}$ に加算していく。 $v_{f,i}$ の値が閾値を超えた時にシーンが切り替わったと判定する。フレーム f がシーンの切り替わりかを判定する式 C_f は以下の(2)のようになる。

$$C_f = \frac{1}{P} \sum_{i=1}^P v_{f,i} > a \quad (2)$$

ここで、 P はダウンサンプリングした際のブロックの数、 a はシーンの切り替えを判定する閾値を示す。ここで、 $P = MN/(mn)$ となる。 a の値の設定により、検出漏れと過検出の間を調整することができる。漫画形式で要約をする際に必要な画像を選ぶという目的から、検出漏れを避けることを優先とする。

過検出を抑制するために、さらに判定条件を加える。平井ら[3]の手法にならない、連続した 2 つのフレームの輝度を比較する判定手法を利用した。画像に Y 信号変



(a) 目が閉じているシーン (b) 目が開いているシーン
エッジピクセル数: 14226 エッジピクセル数: 15442

図 4: 目の描写によるエッジピクセル数の差

©サンライズ・ラブライブ! School idol project. バンダイビジュアル/ランティス/ブシロード/アスキー・メディアワークス, 2013.

換を行い、求められた輝度値からヒストグラムを作成し、式(3)によりヒストグラム間の距離 $D(H_f, H_{f-1})$ を計算する。

$$D(H_f, H_{f-1}) = \sum_I \left| \frac{H_f(I) - H_{f-1}(I)}{H_f(I) + H_{f-1}(I)} \right| \quad (3)$$

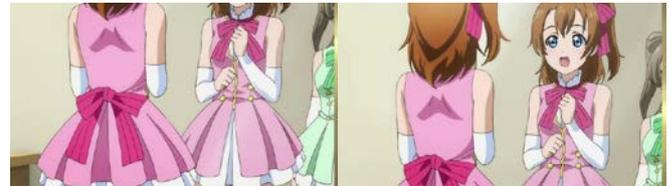
式(3)により求めた距離 $D(H_f, H_{f-1})$ が閾値を超えたときシーンの変更を判定する。ここで $f, f-1$ は隣接フレーム番号、 $H_f(I), H_{f-1}(I)$ は輝度値 I の頻度を示す。上記の式(2)(3)で求めた値がそれぞれ設定した閾値を超えたときに、シーンの変更を判定するようにした。これにより色差によるシーンの検出を行うつつ過検出を抑えることに成功した。式(2)(3)を組み合わせた際の値の変化を図 3 に示す。グラフにおける縦軸は計算により求められる値、横軸はフレーム数を示す。赤い点線はそれぞれの閾値を示し、閾値を超えシーンが変更されたと判定されたフレームには赤い丸を付けている。図 3(b)の青い丸は、二つの手法を組み合わせたことにより過検出が抑制されたフレームを示す。

しかしながら、フィルムコミックの特徴で 2 コマに渡りキャラクターの動きや、長いセリフを表現することが多いというものがある。式(2)(3)の判定をそのまま組み合わせただけでは、キャラクターの動きを別々のキーフレームとして判定することはできない。そのため、2つの判定法に意味を持たせることとした。式(2)で判定されるシーン変更はブロックごとに色差を比べるため、物体の大きな動きに対しては過検出が起きやすい条件となっているので、画面内での物体の動きによる変化を判定させる。式(3)で求められるシーン変更判定では、Y 信号変換したヒストグラムを比較するため、物体の動きに対しては比較的過検出をしないようになっている。これらの性質を利用し、式(2)のみ閾値を超えた場合、キャラクターの大きな動きとし 1 枚までのみ判定を行う。式(2)(3)が同時に閾値を超えたときに、シーンの変更点と判定する。よって同一ショット内でキャラクターなどが大きく動く場合は合計 2 枚のシーンを判定する。この際、式(2)の閾値 a は図 3(b)の通常の閾値の点線上部の、間隔短い赤い点線のような高い値とし、通常よりも過検出を抑える設定とする。



図 5: フェードインのシーンにおけるキーフレーム判定例

©サンライズ・ラブライブ! School idol project. バンダイビジュアル/ランティス/ブシロード/アスキー・メディアワークス, 2013.



(a) 肌色検出なし (b) 肌色検出あり

図 6: 肌色検出の結果

©サンライズ・ラブライブ! School idol project. バンダイビジュアル/ランティス/ブシロード/アスキー・メディアワークス, 2013.

4.2 キーフレーム検出

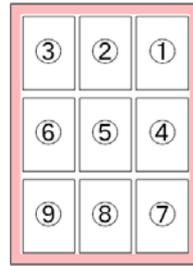
見栄えのするフレームを検出するためのキーフレーム判定の手法について説明する。本研究でいうキーフレームとは、CG アニメーションのキーフレーム法におけるものではなく、フィルムコミックのコマに相応しいと判定して抽出したフレームと定義する。キーフレーム検出にはエッジ検出・肌色検出・音声検出を用いた。これら 3つのキーフレーム検出に重みを加えた特徴量を比較し、検出を行う。

最初に、エッジ検出について説明する。見栄えの良いシーンを抽出する際に考えられるのが、キャラクターが目を開いているかどうかである。動画は画像の連番であるため、目を閉じているシーンから開いているシーンまでに、キーフレームとしては好ましくない”中割り”に該当するフレームも存在する。本研究では、目の描写の複雑性に着目し、閉じている場合と開いている場合でエッジ数がどう変化するかを実験した。それらの画像をキャニー法によりエッジ抽出し、ピクセル数を計算した結果、図 4(a)、(b)のように検出したキャラクターが目を開いているシーンは、エッジのピクセル数が多いことが分かった。実験の結果から、フレームの画像のエッジと判定されたピクセル数と比較し、ピクセル数が多いフレームをキーフレームとする。また、図 5 のような、フェードイン・フェードアウトで微量な色の変化が起きていく場合、フェードシーンより、フェード後のフレームの方がエッジのピクセル数が多いため、上記のエッジ検出により、フェードシーンがある場合でも、図 5 の赤枠に囲まれたような確かなシーンを検出することができる。

二番目に、肌色検出について説明する。エッジ検出のみでキーフレームの検出を行うと、カット内で服など顔以外のエッジが複雑な場合、顔が映るフレームよりも、エッジの複雑な対象が多く映るシーンが検出されてしまう。そこで、カット内で肌色領域と思われる色相 H が $0 \sim 44$ 度の範囲のピクセル数を比較する。色相 H の範囲に関しては効果や肌の色によって黄色みがかかることがあるため広めに設定した。しかしながら、同一ショット内で肌色領域が大きく増加する場合は、単純に顔領域の肌色による影響ではなく、背景等の影



(a) 断ち切りゴマの例 (b) 断ち切りゴマを含むレイアウト



	①	②	③	④	⑤	⑥	⑦	⑧	⑨
0,180	右	×	左	右	×	左	右	×	左
30,60	右			右	×	左	下		左
210,240	上	上	上	右	×	左	下	下	下
90,270	上	上	上	×	×	×	下	下	下
120,150	上	上	左	右	×	左	右	下	下
300,330			上				下		

(a) コマの位置番号 (b) ベクトル角度とコマ番号による断ち切り方向

図 7: 漫画における断ち切りゴマの例

図 8: ベクトル角度と断ち切りゴマの対応付け

©冬川基.とある科学の超電磁砲第6巻.株式会社アスキー・メディアワークス.2011.

響が大きいと考えられるため、例外処理を行う。同一ショット内での比較の際、キーフレーム候補として判定されているフレームの肌色領域ピクセル数 p_{max} が、現在フレームの肌色領域ピクセル数である p_f の 10% 以下の場合、次のシーン変更判定時まで肌色領域の重みを 0 とする。この判定を用いると肌色領域の大きい顔の領域など、より見栄えの良いシーンを検出することができる。肌色検出の有無により選択されたキーフレーム結果の違いを図 6 に示す。この判定を用いると肌色領域の大きい顔の領域など、より見栄えの良いシーンを検出することができる。

最後に音声検出について説明する。事前に人間の声を自動抽出した音声ファイルから各フレームにおける音量 V_f を以下の式から計算する。

$$V_f = \sqrt{\frac{1}{S} \sum_{i=0}^{S-1} A_{f,i}^2} \quad (4)$$

ここで、 f はフレーム番号、 S はフレームごとの音波のサンプリング数、 $A_{f,i}$ は f における i 番目のサンプルの強度を示す。この音量は 4.5 節に示すコマとのマッチングにおいても利用する。

以上の計算によって得られたエッジ検出、肌色検出、音声検出のそれぞれの値に重み付けを行った特徴量 S_f を計算し、キーフレームを決定する。特徴量 S_f を求める式を下記に示す。

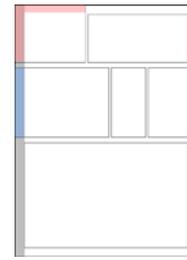
$$S_f = \alpha E_f + \beta V_f + \gamma F_f \quad (5)$$

ここで、 f はフレーム番号、 E_f はエッジのピクセル数、 V_f は音量、 F_f は肌色領域のピクセル数、 α 、 β 、 γ は重み係数を示す。重み付けの処理は通常の検出と、肌色検出の例外処理の 2 通り存在する。分割された各シーン内で特徴量の最大となるフレームをキーフレームとする。

4.3 カメラワーク検出

隣接フレームからオプティカルフローを計算し、求めたベクトルと x 軸のなす角度から、ティルトが使用されているシーン（以下、ティルトシーン）と断ち切りゴマの判定を行う。判定結果はコマとのマッチングとレイアウトに利用する。なお、横方向のカメラワークであるパンのシーンも存在するが、フィルムコミックではパンによるシーンの表現は少ない。また非常に横長になる場合、連結したシーンに存在する人物等が小さくなってしまふ。そのため、パンによる横長のシーンの検出においては本研究では対象外とする。

図 1 右側のようなティルトシーンを 1 つのコマとして扱うための手法を説明する。オプティカルフローから求めたベクトルと x 軸のなす角度が 90 度、270 度の



(a) レイアウト図



(b) 120 度の断ち切りゴマ



(c) 180 度の断ち切りゴマ

図 9: 断ち切りゴマを含むレイアウトの生成例

時に、隣接フレームを利用し、テンプレートマッチングを行う。ティルトシーンでは、フレームごとに共通する部分が存在する。そのため、オプティカルフローから求めたベクトルが x 軸と 90 度をなす場合はフレームの下部をテンプレートに、270 度をなす場合はフレームの上部をテンプレートとする。テンプレートとマッチング対象部分の HSV 表色系における色相 H を比較し、色相 H が 10 度以上異なる画素を数える。画素数が閾値を超えない場合ティルトが行われていると判定し、該当するすべてのキーフレームとし、コマとのマッチングに利用する。

漫画技法として図 7 の (a) のような、断ち切りゴマという描写方法がある。これは (b) のピンク色で示す、印刷範囲の限界までコマを描く技法で①、③のコマが断ち切りゴマである。断ち切りゴマは、軽快な視点移動を行ったり、効果線も合わせてコマに疾走感や迫力を持たせたりすることができる。この技法を使用するかどうか、漫画の作者の裁量に委ねられ、明確に行わなければならないという制約は存在しない。そのため、本研究では「オプティカルフローによるベクトルを持つフレーム」と「レイアウト時のコマの大きさ」を考慮し、視線移動を軽快にするために断ち切りゴマでの描写を行うこととする。断ち切りゴマの判定には、オプティカルフローから求めた x 軸とベクトルのなす角度を利用する。また、ベクトルの角度とコマの位置によっては断ち切りゴマとしない場合が存在する。それらの対応付けのためのコマの位置番号と、ベクトルのなす角度との対応付けを図 8 の (a)、(b) に示す。

漫画の原稿用紙には製本する際に閉じられてしまふノドと呼ばれる領域がある。これは左右どちらのページあたるかで逆転し漫画の作画としては、絵やセリフを書くことは推奨されない。しかしながら、フィルムコミックではノドの領域まで描写することが多いため、本手法ではノドの方向に限らず、断ち切りゴマによる描写を行う。

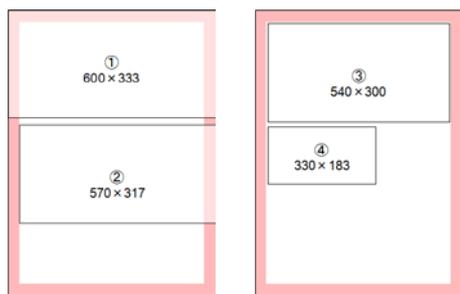
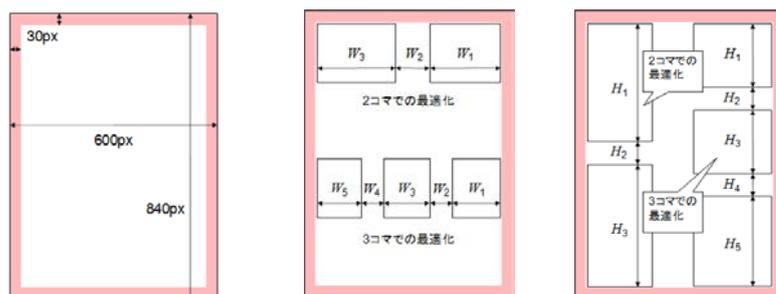


図 10: リサイズの種類



(a) キャンバスのサイズ (b) 横幅領域の定義 (c) 縦幅領域の定義

図 11: キャンバスサイズとコマ領域の定義

表 1 推奨領域とリサイズの対応付け

#1	推奨領域の縦横が元画像の 90%以上かつ、音量が閾値以上
#2	推奨領域の縦横が元画像の 90%以上かつ、ベクトル角度が $0\sim 60^\circ$ 、 $120\sim 240^\circ$ 、 $300\sim 360^\circ$
#3	推奨領域の縦横が元画像の 90%以上もしくは、顔領域の縦横が 60%以下
#4	推奨領域が上記以外の場合

図 9 の(a)に断ち切りゴマを含むレイアウトの生成を示す。(b)、(c)がベクトルの角度により断ち切りゴマとして描写されたコマである。(a)の赤色で塗りつぶされた領域が断ち切りゴマにより拡張された領域部分である。(a)の青色の領域はサイズによる断ち切りゴマであるため、詳細は 4.6 節にて説明を行う。

4.4 推奨領域の計算

コマの推奨領域の検出には顔検出を利用した。推奨領域の検出によりコマのサイズや形を決定する。主に漫画、フィルムコミック共にコマの中心となるのはキャラクターであり、そのキャラクターの顔がコマに配置されることが多い。そのため、キーフレームとして保存された画像から顔検出を行う。顔検出には高山ら [7]の手法の一部を利用した。

高山らの手法では、肌色以外の領域を分割し、それぞれの領域の色をクラスタリングし、対称になると見込まれる目の領域を用いて判定を行う。しかしながら、この手法は理想的な正面を向いたキャラクターに対して有効な手法である。実際のアニメ作品では必ずしも目の領域が存在しているわけではなく、横顔などで目の領域が 1 つしか存在しない場合もある。本手法では、横顔などもコマに入れることを考えるため、目の領域による検出を行わないこととした。これらにより検出された顔領域をすべて含む領域を、人物が存在する領域とし、その領域をコマの大きさを決める際の推奨領域とする。

4.5 キーフレームのマッチング

コマとのマッチングにおいてはコマのレイアウト処理を行う際の、コマの推奨領域とのマッチングを行う。はじめにサイズとのマッチングを行う。4.4 節の顔領域検出から得られた推奨領域の割合から画像をリサイズする。リサイズするサイズの種類を図 10、領域の割合とリサイズの対応を表 1 に示す。

サイズとのマッチングを行った後、#4 にマッチングされた画像に限り、得られた推奨領域をコマのレイアウト処理で利用する。#1、#2、#3 にマッチングされた画像の推奨領域はリサイズされたサイズとなり、直接コマの配置処理に利用される。なお、ティルトシーンに限っては横幅のサイズを #2 とし、縦幅はティルトシ

ーンの大きさによって可変となる。

4.6 レイアウトの最適化

コマのレイアウト処理にはコマとのマッチングで求めた縦横の幅の推奨領域を利用した、最適化処理を行う。レイアウト処理においての最適化は推奨領域を理想の領域として計算を行う。図 11 の(a)に示すように、レイアウトを描写するキャンバス（原稿用紙）のサイズを 600×840 画素とする。そこから上下左右 30 画素を断ち切り領域とする。

レイアウト処理においてはコマとのマッチングにより得られた推奨領域からコマの横幅、縦幅を計算するが、マッチングにより表 1 に示す #1、#2、#3 のコマについては特定条件以外では横幅の計算は行わず、#4 のコマに関してのみページに対して製本時に閉じられるノドの領域とは逆側を断ち切りゴマとするように配置処理を行う。すなわち、右ページであれば、左側がノドにあたるため右方向に断ち切りを行う。図 11 の(b)、(c)に示すように、並ぶコマの数が 2 つか 3 つかで最適化する領域が異なる。断ち切り幅を考慮した p_w 、 p_h は以下の式で計算する。

$$p_w = 600 - (\mu_0 + \mu_1) \quad (6)$$

$$p_h = 840 - (\mu_2 + \mu_3) \quad (7)$$

以上の条件を踏まえ、次の式のようにコマやコマ間の横幅の領域の最適化を行う。

$$\arg \min_w \sum_{i=1}^n (W_i - S_i)^2 \quad (8)$$

$$W = \{W_1, W_2, \dots, W_n\}$$

ここで、 W_i は i 個目のコマまたはコマ間の横幅、 S_i は i 個目のコマまたはコマ間の推奨領域の横幅を示す。最適化を図る変数の数 n は図 11 の(b)に示すように 3 または 5 になる。同様の計算で、以下の式からコマやコマ間の縦幅の最適化を行う。

$$\arg \min_H \sum_{i=1}^n (H_i - S_i)^2 \quad (9)$$

$$H = \{H_1, H_2, \dots, H_n\}$$

ここで、 H_i は i 個目のコマまたはコマ間の縦幅、 S_i は i 段目の横に並ぶコマまたはコマ間の推奨領域の縦幅の平均を示す。 n は図 11 の(c)に示すように 3 または 5 になる。最適化の方法は最小二乗法を用いた。また、

W_i や H_i に重みを与えることで、コマ幅の精度を調整することができる。漫画やフィルムコミックではコマ間の幅が安定した幅を持つことが望まれるため、コマ間の幅を求める式に重みを与え、一定の幅を保持できるようにした。

5. 実験と評価

5.1 実験条件

提案手法の実装には C++、OpenCV を用いた。実験環境として Intel Core™ i7-3820 CPU 3.60GHz、8.00GB RAM 搭載の PC で実験を行った。実験にはアニメ作品はプロダクション I.G 制作の「東のエデン劇場版 I The King of Eden」(81 分) を利用した。それぞれオープニングとエンディングは事前にカットし、本編のみ(約 74 分)の状態とした。また利用した動画は、解像度 720×480 、29.97fps である。そのままだと、フレーム補間による同一フレームを含む 134,297 枚のフレームを比較していくことになるため、実験では 4 分の 1 の 33,574 枚のフレームを利用した。実験で使用したパラメータを表 2 と表 3 に示す。

5.2 結果

実験の結果 855 枚のキーフレームを検出した。利用したフレームは 33,574 枚であるため約 40 分の 1 に映像要約することができた。キーフレームの検出例を図 12、ティルトシーンの連結結果例を図 13 に示す。

これらのキーフレームを利用し、レイアウトを行った例を図 14(1-a)と(2-a)、(3-a)に示す。実験においては、評価を考慮してフィルムコミックを参照し、本研究で対応した漫画技法が含まれるシーンをいくつか選択しレイアウトの生成を行った。

5.3 評価

キーフレーム検出とレイアウトについて評価する。キーフレーム検出では、検出したキーフレームの検出枚数を評価する。レイアウトでは、フィルムコミックとレイアウト画像を比較する。比較するフィルムコミックには「フィルムコミック 東のエデン劇場版」を使用した。

表 2 パラメータ

ダウンサンプリング	$M = 720, N = 480$ $m = 20, n = 20$
シーン変更判定	$\Delta_{min} = 40, \Delta_{max} = 60,$ $a = 0.038$ (色差のみの場合 0.06) D の閾値 $\theta = 50.0$
キーフレーム検出	$S = 2943, \alpha = 0.3, \beta = 1.0, \gamma = 0.1$
コマ間の推奨領域	横幅: 12px, 縦幅: 15px

表 3 推奨領域とマージンの関係

	μ_0	μ_1	μ_2	μ_3
右ページ: (推奨領域の幅の合計) > 600	0	0	30	30
右ページ: (推奨領域の幅の合計) > 570	0	30	30	30
左ページ: (推奨領域の幅の合計) > 600	0	0	30	30
左ページ: (推奨領域の幅の合計) > 570	30	0	30	30
(推奨領域の高さの合計) > 840	30	30	0	0
(推奨領域の高さの合計) > 810	30	30	0	30

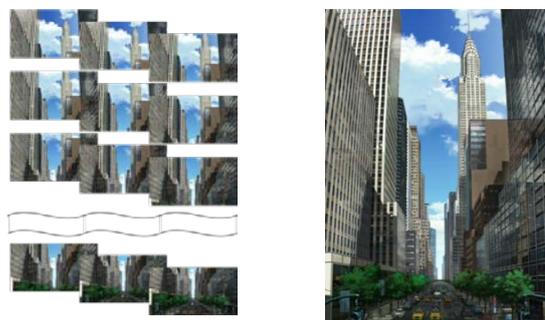
表 4 キーフレーム要約フレーム数比較

	本手法	フィルムコミック
要約フレーム数	836 枚	862 枚



©プロダクション I.G. 東のエデン劇場版 I The King of Eden. アスミック/フジテレビ. 2010.

図 12: キーフレーム検出例。各行は同一シーンにおいて左) 検出し、中央) エッジ検出、右) 音声検出による結果。



©プロダクション I.G. 東のエデン劇場版 I The King of Eden. アスミック/フジテレビ. 2010.

図 13: ティルトシーンの連結結果例

はじめにキーフレームの検出枚数をフィルムコミックで利用されたフレーム数と比較を表 4 に示す。検出枚数はキーフレーム検出の方法に関係なく同一の枚数となる。次に、本手法によるレイアウト画像を、フィルムコミックのレイアウトを再現した画像(図 14(1-b)と(2-b)、(3-b))と比較した。この結果に示すようにフィルムコミックに近いレイアウトやティルトのコマを再現することができた。

検出されたキーフレームがフィルムコミックで利用されたキーフレームと完全に一致するかを評価したものを表 5 に示す。ここで言う「完全に一致」とは、中割りによるフレームの数ピクセルのズレも許容しない厳密な抽出評価である。

同一キーフレーム数の一致率において、エッジ+音声検出が最も精度が高く、音声検出もそれに続くことから、フィルムコミックにおいては音量を重視したフレームを利用する傾向があると考えられる。キーフレームの検出を行わない場合の一致率が最も低いことから、キーフレーム検出は有効であると考えられる。

入力アニメ内のティルトシーンの数と、提案法によって検出できたティルトシーンの数、またフィルムコミックで大きなコマとして使用されている数を表 6 に示す。77%の検出率が実現できており、フィルムコミックで使用されるコマ数を十分包括できることがわかる。

レイアウト生成の失敗例を図 15 に示す。本手法ではフレームの全体を利用しつつも小さいコマとする処理には対応していないため、フィルムコミックで縮小処理が行われている場合はレイアウトで誤差が出てし



(1-a)

(1-b)



(2-a)

(2-b)



(3-a)

(3-b)

©プロダクション I.G. 東のエデン劇場版 I The King of Eden. アスミック/フジテレビ. 2010.

図 14: レイアウト画像比較。それぞれ a は提案法によるレイアウト、b は実際のフィルムコミックのレイアウト。

まう。また、映像効果やシーンによっては、キャラクターが存在していても顔検出で一切検出ができないフレームも存在し、大きな誤差の原因になると考えられる。キーフレームとしては検出できているため、ユーザーの意図に合わない場合は、ユーザーの介入によってフレームレイアウトを調整する方法を検討している。

表 5 検出されたキーフレームの完全一致率
フィルムコミックに使用されている 862 枚のフレームとの比率

	本手法による抽出枚数	一致率
検出なし	300 枚	34.8%
エッジ検出	395 枚	45.8%
音声検出	415 枚	48.1%
エッジ+音声検出	419 枚	48.6%



(a)

(b)

提案法によるレイアウト

実際のフィルムコミック

©プロダクション I.G. 東のエデン劇場版 I The King of Eden. アスミック/フジテレビ. 2010.

図 15: 実際のフィルムコミックとレイアウトに大きな差異がある例

表 6 ティルトシーンの検出結果

アニメ作品内の ティルトシーンの数	提案法によって 検出できた数	フィルムコミック内の 縦長のコマ数
13	10	7

6. まとめと今後の課題

本研究では、フィルムコミックに着目し、アニメ作品の映像を入力することで、コマに必要なキーフレームの検出からの漫画形式の映像要約を目的とした手法の提案を行った。画像から得られる、色特微量やエッジ検出、オプティカルフローに加え、音声のある動画であることに着目し、音声検出を考慮することで、漫画のコマに相応しいキーフレームを検出することに成功した。また、得られたキーフレームから、コマとのマッチングと最小二乗法によるレイアウト処理を行い、漫画技法を適用した漫画形式のレイアウト画像の出力に成功した。

コマのレイアウト処理では、フィルムコミックの際に必要な吹き出しの長さや領域を考慮していないため、音声ファイルの利用を音量によるキーフレームの決定だけでなく、セリフ部分や長さの検出から吹き出しの大きさを推定し、コマ割りに利用する処理などが必要であると考えられる。

文 献

- [1] Kurihara Kasutaka, "Cinema-Gazer: a System for Watching Videos at Very High Speed", In Proceedings of AVI'12, pp.108-115. 2012.
- [2] Naomichi Murakami, Eisuke Ito, "Emotional video ranking based on user comments", iiWAS '11 Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services pp.499-502. 2011.
- [3] 平井辰典, 中野倫靖, 後藤真孝, 森島繁生. シーンの連続性と顔類似度に基づく動画のコンテンツ中の同一人物登場シーンの同定. 映像情報メディア学会誌 Vol.66, No.7, pp.J251-J259. 2012.
- [4] Ying Cao, Antoni B. Chan, Rynson W. H. Lau, "Automatic Stylistic Manga Layout", In Proceedings of SIGGRAPH Asia 2012, Vol. 30 (6), 141:1-141:10, 2012.

- [5] 福里司, 平井辰典, 大矢隼士, 森島繁生. アニメ作品のキーフレーム検出による漫画形式の映像要約手法. VC/GCAD 合同シンポジウム 2013, No.19, 2013.
- [6] 福里司, 平井辰典, 大矢隼士, 森島繁生. アニメ作品におけるキーフレーム自動抽出に基づく映像要約手法の提案. 画像電子学会誌ビジュアルコンピューティング論文特集号, Vol.42, No.4, p.448-456, 2013.
- [7] 高山耕平, ヘンリージョハン, 西田友是. 特徴量抽出によるアニメキャラクターの顔認識とキャラクター検索. VC/GCAD 合同シンポジウム 2012, No.2, 2012.
- [8] 増田太郎, 福里司, 平井辰典, 大矢隼士, 森島繁生. 顔領域の画像類似度に基づく動画中のキャラクター登場シーン推薦. VC/GCAD 合同シンポジウム 2013, No.45, 2013.