

## オマージュ作品とのシーンマッチングを用いた ビデオクリップのハイライトシーン抽出

牛越達也 河野恭之

映像作品のコミック形式への変換を目指し、鑑賞者の興味を集中する場面を抽出し、興味度に応じて強調した画像を提示するシステムを作成し、一定の評価を得た。web 上で公開されている複数のオマージュ作品とオリジナル作品と一致する区間を検出することで、二次創作者の興味が集まっている場面を特定、抽出する。

### 1. はじめに

本研究では、映像作品のコミック表現変換を目指し、オリジナル作品と複数のオマージュ作品と一致する区間を検出することでユーザの興味情報に即した重要度を設定する。オリジナル作品から重要シーンの画像を抽出して、重要度に強調した画像をマンガ表現生成の前段階の表現物としてユーザに提示するシステムの構築、検証を行った。近年、インターネットの発達と共に、映像、音楽などマルチメディアコンテンツの web 上での配信、共有が盛んに行われている。そのため、プロが作成したコンテンツだけでなく、個人が撮影、又は作成したコンテンツである UGC も爆発的に増加しており、そのようなコンテンツの配信、管理、検索方法の問題が顕在化している。中でも映像コンテンツは、Youtube やニコニコ動画といった動画共有サイトで自由に閲覧できる環境が整っていることもあって個人化された利用が増加しており、同時にタグなどコンテキストベースの情報分類も進んでいる。しかし、このような情報の分類もまだ不十分であり、特に映像コンテンツの視聴には実時間が必要であることから、検索効率が悪い。そのため、映像コンテンツの検索の効率化のためには、映像コンテンツの大スジを事前に把握するような手法が必要である。

そうした表現にマンガがある。映像をマンガ的にマップする研究はこれまでいくつか行われており、一定の有効性が示されている。マンガ的手法を用いて、ユーザが映像を能動的に視聴しメディア内容の理解することを支援する試みで内橋は、ビデオ内容を大

きさの異なるキーフレームを使って視覚的に捉えやすい形態で表示するビデオ要約を用いて、インタラクティブにビデオ再生を行うインタフェースを提案している[1]。この研究では、連続したイメージを持つフレーム区間をセグメントと定義し、セグメントをクラスタリングすることで、同じクラスに属するセグメントを類似セグメントとしている。その類似セグメント量をキーフレームの配置、強調の根拠としてマンガ調の表現を生成している。しかし映像表現を考えるなら、類似した場面が多いから視聴者がそこに引きつけられるのではなく、たとえ短い場面でもその場面が語る物語に視聴者は引きつけられるものであり、本方式による強調が適切でない対象は少なくないと考えられる。したがって、強調の根拠としては弱いものになっている。中村らは、ニコニコ動画のシーンごとのコメント投稿数に着目し、コメント投稿数に応じて各シーンのキーフレームをマンガ的手法で提供するインタフェースを構築している[2]。彼らは、シーンごとのコメント投稿数のユーザの印象に残ったシーンと一致すると仮定し、各シーンに対するコメント投稿量を強調の根拠としたマンガ調の表現を生成している。しかし、コメント数がユーザの印象に繋がるというアイディアは斬新だが、コメントがコメントを助長する例など、コメント事態がユーザの印象に影響を与えていることが問題となっている。そこで本研究では、オリジナル作品とオマージュ作品の一致区間を検出することで鑑賞者の興味を集中する場面を抽出し、鑑賞者の興味情報に即して重要度の設定を行う。オマージュ作品とは、自身の受けた感動を他者と共有する目的で私的に編集されたものであり、オリジナル作品に対する鑑賞者個人の感想も含めた、個人の興味情報が含まれている可能性が高い。オマージュ作品は動画共有サイトにおいて、複数のユーザによってタグ付け分類されており、人気のあるオリジナル作品にはより多くのオマージュ作品が登録されている。また、動画共有サイトで総再生回数などによってランキングされており、質の良い鑑賞者の興味情報を得ることができる。そのため、複数のオマージュ作品に採用されているオリジナル作品の断片を集めることで、オリジナル作品を「愛する」鑑賞者たちの思考、視点を採り入れた重要度の設定が可能となる。

### 2. 概要

本研究では、動画作品の構造を解析する。ここで、映像作品の構造を 3 つの概念にわけ、これらの概念による Video 構造を図 1 に示す。Frame とは Video を構成する各画像、Cut とは連続するイメージを持つ Frame 区間、Scene とは意味的に一致する時系列に並んだ



図1 Video構造の定義

Cut 区間である。一般に、同一 Cut に属する Frame 画像では、背景、対象ともに大きく類似している。逆に、Cut 境界の前後ではこれらは大きく異なる性質を持つ。Scene は、共通の対象を写す、似たような場面構成が続くといった、場面ごとの最小単位に連続な Cut をまとめたものである。また、ある場面を軸に他の場面が映し出されることが続くといった共起関係が見られる Cut 区間も一つの Scene とする。本研究では、こうした性質を利用して Cut の分割、Scene の決定を行い、その解析結果を基にオマージュ作品を用いたハイライト検出を行う。以下に全体の処理の流れを示す。

#### Step1. Cut 検出

RGB 色ヒストグラムを用いて、オマージュ作品、オリジナル作品を Cut に分割する [3]。

#### Step2. Scene の検出

オリジナル作品の Cut ごとの平均ヒストグラムを用いて、クラスタ数を変えながらクラスタリングを行う。クラスタリングには K-means 法を用いる。クラスタ数を変えながら得た結果を統合し、各 Cut の結合度を算出する。隣接する Cut 間の結合度が閾値よりも下まわり、次の隣接する Cut 間で閾値を上回るなら、Scene 境界とし、オリジナル作品を Scene に分割する [4]。

#### Step3. フォークソノミー方式によるハイライト検出

オリジナルの Cut とオマージュの Cut を総当りで類似度判定する。類似度の評価には、Frame の動きを同期させながら、色相、グレースケール画像のピラミッド構造を作成し、多重解像度解析を行う。Step1 から Step3 までの処理を複数のオマージュ作品に適用し、結果を統合する。

#### Step4. 重要度の設定

オリジナル作品の Cut に対する一致判定の集中度と、Scene の中で採用されたオリジナル作品の Cut の集中度を評価し、各 Cut と Scene に重要度の設定を行う。次章より、フォークソノミー方式によるハイライト検出の詳細を説明する。

### 3. フォークソノミー方式によるハイライト検出法

#### 3.1 フォークソノミー

フォークソノミーとは、folks(民衆)と taxonomy(分類)をあわせた造語であり、ウェブサイト上の価値ある情報に複数の利用者自らが複数の tag を付け加えることで検索できるように分類していく方法のことである。本研究では、前述したとおり複数のオマージュ作品とオリジナル作品との Cut の一致判定を行い、ユーザの注目が集中する Cut をハイライトと判断する。オマージュ作品は、自分の受けた印象や感動を他者と共有する目的でオリジナル作品を私的に編集したものである。そのため、オマージュ作品に用いられている Cut はオマージュ編集者の印象深かった場面である可能性が高い。また、複数のオマージュ作品と比べることで、複数のユーザが同様に強い印象を受けた Cut を特定することができる。また、オマージュ作品は動画共有サイトでオリジナル作品のタイトル、敬称、愛称といった tag によって容易に収集することができる。そのため、本研究でのハイライト検出法をフォークソノミーによるハイライト検出と呼ぶこととする。ハイライト検出を行うために、以下のポイントが必要である。

- ・オマージュ作品の中で用いられるオリジナル作品の断片の検出
- ・複数のオマージュに用いられているオリジナル作品の場面の検出

次節で、その手法を述べる。

#### 3.2 ハイライト検出

オリジナル作品と各オマージュ作品を互いに Cut に分割し、オマージュとオリジナルの各 Cut を総当りで判定する。Cut 間では、Frame の動きを同期させながら、各 Frame の色彩情報と形状情報を基に一致判定し、複数のオマージュ作品より一致(以後、Link と述べる)が得られた Cut をハイライトとする。(図2) 以下にアルゴリズムを示す。

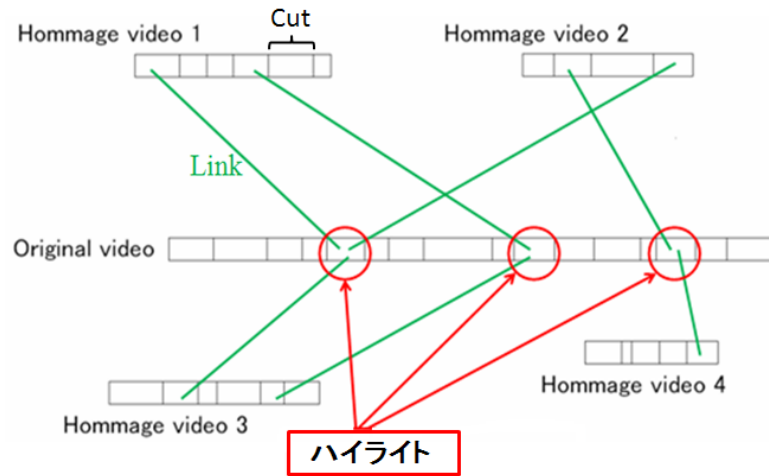


図2 フォークソノミーによるハイライト検出

オリジナルの Cut :  $C_m$  ( $m \in [1, M]$ )

オマージュの Cut :  $C_n$  ( $n \in [1, N]$ )

$C_m$  の始点 :  $S_m$

$C_n$  の始点 :  $S_n$

$C_m$  の長さ :  $L(C_m)$

$C_n$  の長さ :  $L(C_n)$

**Step1.**  $C_m$  と  $C_n$  の, Cut 幅の差  $D$  ( $D = L(C_m) - L(C_n)$ ) を計算する.

$D \geq 0$  なら **Step2** に進む. そうでないなら, **Step4** に進む.

**Step2.**  $C_n$  の始点を  $S_n + d$  ( $d \in [0, D]$ ) として,  $C_m$  と並行する Frame の RGB 差分の和を求め, その和が閾値以下であるなら  $C_m$  と  $C_n$  は一致する Cut として, **Step5** に進む. そうでないなら, **Step3** に進む.

**Step3.**  $d = d + \text{step}$  ( $\text{step}$  : Frame 幅) とする,  $d > D$  となるなら, **Step4** に進む. そうでないなら, **Step2** に戻る.

**Step4.**  $n = n + 1$  とする.  $n > N$  となるなら,  $n = 1$  として **Step5** に進む. そうでないなら, **Step1** に戻る.

**Step5.**  $m = m + 1$  とする.  $m > M$  となるなら, 処理を終わる. そうでないなら, **Step1** に戻る.

**Step2** の Frame ごとの RGB 差分の計算では, オリジナル, オマージュ両 Frame よりグレースケール画像, HSV 色空間を用いた H 画像を作成し, それぞれの画像を  $16 \times 16$  画素領域ごとに量子化することで, 互いに 2 段階のピラミッド構造を作成する. グレースケールの低い解像度の画像は, 原画像の粗い形状情報を持つものであり, H 画像の低い解像度の画像は粗い色相情報を持つものである. 互いの Frame の比較では, はじめに低い解像度のグレースケール画像と H 画像の比較を行い, 互いに近い値を持つことが確認できた領域に対して, 原画像同士の RGB 差分の計算を行う. 以下にアルゴリズムを示す.

入力画像 1 :  $I$

入力画像 2 :  $I'$

入力画像 1 の RGB 値 :  $(R, G, B)$

入力画像 2 の RGB 値 :  $(R', G', B')$

**Step2\_1.** 各入力画像よりグレースケール画像と HSV 色空間の H 画像を作成し,  $16 \times 16$  画素領域ごとに量子化したモザイク画像を作成する.  $I$  のグレースケール画像より作成したモザイク画を  $P$ , HSV 色空間の H 画像より作成したモザイク画を  $Q$  とし,  $I'$  より同様に作成されたモザイク画をそれぞれ  $P'$ ,  $Q'$  とする.

**Step2\_2.**  $P$  と  $P'$ ,  $Q$  と  $Q'$  の同位置の画素に対して, 差分  $D, D'$  を計算する. **Step2\_3** に進む.

**Step2\_3.** ( $D < T$  &  $D' < T'$ ) ( $T, T'$  : 閾値) なら **Step2\_4** に進む. そうでないなら, **Step2\_2** に戻る.

**Step2\_4.**  $P$  と  $P'$ ,  $Q$  と  $Q'$  の画素に対応する  $I$  と  $I'$  の領域に含まれる画素に対して, RGB 差分二乗和を計算する. 式は以下である.

$$Dis = Dis + (R-R') \times (R-R') + (G-G') \times (G-G') + (B-B') \times (B-B')$$

P と P', Q と Q' に探索すべき画素が残っているのなら, **Step2\_2** に戻る. 残っていないなら **Step2\_5** に進む.

**Step2\_5.** 以下の計算を行う.

$$Dis = \text{sqrt}(Dis / TotalPixel)$$

(TotalPixel: RGB 差分二乗の計算を行った画素数)

Dis の出力を行って, 処理を終わる.

#### 4. 重要度の設定

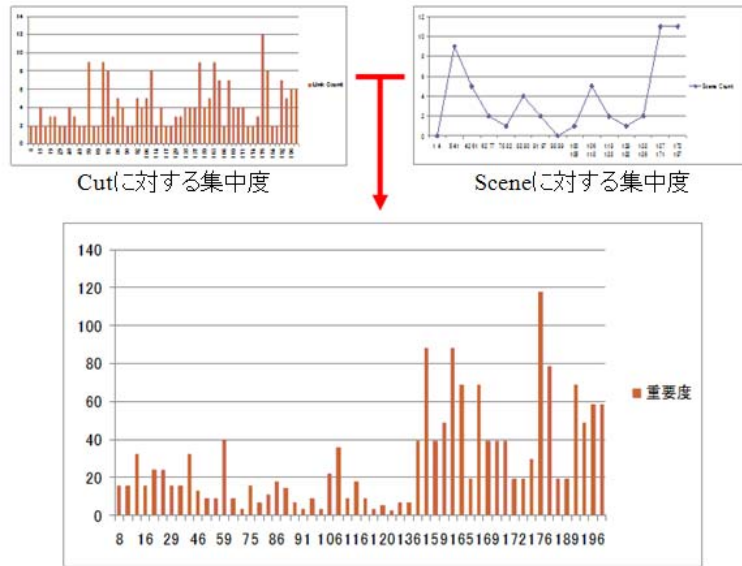


図3 重要度の設定

それぞれ抽出された画像群は, Link 数による Cut に対する興味の集中度と複数の Link を得られた Cut の Scene の中での集中度を定量的に判断することによって, 各 Cut に重要度を設定する. オリジナル作品を Scene に分割し, 複数の Link がはられた Cut の集中度を

各 Scene に対する重みとして算出する.

$S_n$ :  $n$  番目の Scene ( $n \in [1, N]$ )

$M$ : 複数の Link がはられた Cut 数

$G_n$ :  $n$  番目の Scene に対する重み値

$$G_n = \frac{S_n}{M}$$

Scene に対する重みを用いて, 各 Cut の Link 数に応じて重要度を算出する. 図3は, 評価実験に用いたアニメーション映画「秒速5センチメートル」(詳細は, 後述)の各 Cut の Link 数に対して, Scene の重み付けをした結果として設定された各 Cut の重要度である.

$C_m$ :  $m$  番目に採用された Cut ( $m \in [1, M]$ )

$S(C_m)$ :  $C_m$  を含む Scene の番号

$I_m$ :  $m$  番目に採用された Cut の重要度

$$n = S(C_m)$$

$$I_m = C_m \times G_n$$

#### 5. 評価実験

##### 5.1 実験方法

20~30 代の 12 名を被験者にして評価実験を行った. 実験に用いた作品は, アニメーション映画「秒速5センチメートル 第1章 桜花抄」である. ハイライト抽出に用いるオマージュ作品は, ニコニコ動画投稿作品から収集した. 収集基準は, 本作品のタイトルで検索されるものの中で, 2008 年 12 月 30 日までの総再生数の上位 12 本である. 実験内容は, 以下に示す.

1. オリジナルの映像作品を見せる.
2. 前章で述べた処理によって, 抽出された画像群について説明する.
3. 抽出された画像群を見せ, 4 段階評価の質問項目と記述形式のアンケートに答えてもらう.
4. 抽出された画像群 Level 1 との PV(Promotion Video)との比較を行い, アンケートに答えてもらう.

抽出された画像群は, Link 数に応じて 3 段階のレベルに分けて提示する.(図4) 実験

に用いた作品は、12本のオマージュ作品と判定を行ったので Link 数の最大値は 12 である。Link 数が 10 以上の Cut より抽出された画像群が Level 3、同様に Link 数 6 が以上と 2 以上の画像群がそれぞれ Level 2 と Level 1 である。図 3 は、各 Level において抽出されたオリジナル作品中の区間を示したものである。Level 3 では、物語のクライマックス部分結果が結果として抽出されている。それに対して、Level 2 は Level 3 のクライマックス部分に加えて、クライマックスに向かう物語の導入などが Level 3 に加えて抽出されている。Level 1 は、物語全体を網羅するように抽出がされている。提示する画像は、各 Cut の中心の Frame を採用した。

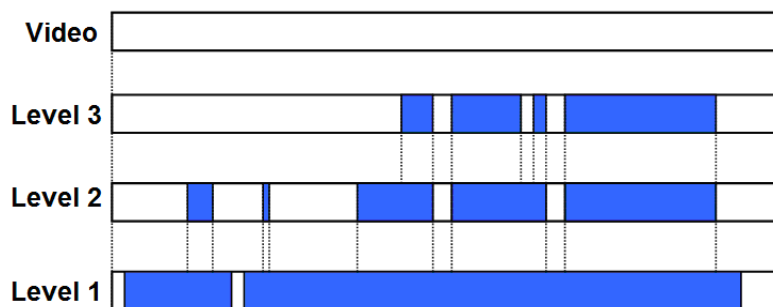


図 4 各 Level の抽出範囲

強調は、4 章で設定した重要度の高い順にソートし、上位 10% までが含まれるものを強調が最も大きいもの、さらに上位 10% を除いた画像の 20% までが含まれるものをそれに準じるものとする。詳細は表 1 に示す。画像の配置は時系列の並びにする。

表 1 強調のルール

強調	サイズ	全体に占める割合
大	320 x 290	10%
中	265 x 200	20%
小	205 x 155	70%
等倍	175 x 130	100%

### 5.1.1 質問項目

質問項目を以下に示す。

• 4 段階評価の質問項目：

1. 画像の流れから、ストーリーが感じられますか？
2. 画像サイズによる強調は、適切ですか？
3. 画像の提示枚数は、適切ですか？

• 記述式の質問項目：

4. 提示画像の中で不要だと感じた画像に、○を打ってください。
5. とれてほしかった画像など、気になった点があったら書いてください。

• PV との比較項目：

6. Level 1 と比べて、内容について受ける印象はどのように変化したか書いてください。

4 段階評価の質問項目は、提示された画像の流れから受ける印象を率直に良、悪で表現するためである。1 の質問は、それぞれの提示された画像群からシナリオを感じ取れる程度を聞いたものである。各 Level の画像群は、それぞれ抽出範囲が異なり、Level 1 以外は全体を網羅するものとはいえないものであるが、物語を部分的に切り出しているものである。その切り出された画像群から感じるシナリオの流れを 4 段階で評価する。2 の質問は、各 Level に設定された重要度を適用して画像を強調した結果受ける印象を聞いたものである。本研究では、Cut に対する Link 数を、抽出された Cut の Scene の中で集中度を重みとして定量的に評価することで重要度を設定した。それによって、強調される画像がユーザにとって印象的に受け取られた場面と一致しているかを評価する。3 の質問は、提示された画像数は切り出された場面を表現するに必要十分であったかを聞いている。抽出された画像群は、Level が下がるごとに画像数が肥大化し、それぞれその画像数に見合った物語を表現している。その表現するに冗長な画像が多いか、画像が足りていないかを評価する。

記述式の質問 4 では、提示された画像の中にどれだけの不要な画像が含まれかを数値化する目的で聞いている。また、5 の質問とあわせて、不要な画像の分類化を行う目的も含んでいる。5 の質問は、提示された画像の中で足りていなかった画像について、画像の提示傾向について聞いている。

PV との比較項目は、本手法によって抽出された部分と製作者側の注目させる意図を持って表現した部分との比較を行うものである。本手法に抽出されるものは、作品の中で視聴者側からみた注目すべき場面である。それにより、表現されるものが作品の製作者側によって表現されるものとの違いを図る目的で、宣伝目的で製作者側によって演出された PV より抽出される画像群との比較を図る。抽出方法は、同様に前章 4.2 の判定を用い、画像の強調は行わない。PV は、物語全体を表現したものであるため、比較対象には Level 1 の画像群を用いるものとする。図 5 は、PV の抽出範囲を示したものである。PV は、Level 1 に含まれる形で抽出されている。



図 5 PV と Level 1 の抽出範囲の比較

## 5.2 評価

実験の結果、Level が下がるごとに全体のストーリーの把握がしやすくなる評価を受けた。各 Level では、それぞれ網羅している領域が大きく異なっており、同時に全体として提示する画像数も増加している。全体のストーリーの把握が向上して行くことは、それにとともに結果が得られたことがいえる。

画像の強調は、4段階評価では Level が下がるごとに強調の精度が上がっていることが確認されたが、全体としては低い評価を受けた。その要因としては、被験者がダイジェストとして不要と考える画像が多く抽出されてしまったことによって、Scene に対する集中度が高められてしまっていることがあげられる。各 Level に共通して、不要とされた画像は 20~30%含まれると判断されている。それら不要とされた画像を記述式の質問項目 5 によせられたコメントから次のように分類する。

### ・背景画像

本作品は、特定のシーンを際立たせる目的や時間の流れを感じさせるために、情景描写が多用されている。その傾向は、本作品を用いて表現されているオマージュ作品に

もみることができるため風景の画像が多く抽出される結果となった。

### ・冗長と感じられる画像

#### ・同じ画像

映像の表現として同じ対象を写した Cut を軸に、他の Cut を写すことを繰り返す映像表現が影響している。

#### ・類似した画像

対象も場面も同じであるがカメラワークが異なるものである。

これらの画像群は、映像表現と物語を画像でマップした表現との表現形態の違いによって不要とされたものである。これらのような、不要な画像によって Scene の中での集中度が高められてしまうことが強調失敗の要因である。

その他に強調を失敗した要因としては、Link 数じたいが不足した場合がある。本実験で強調を失敗した画像では、作品の中でも話が盛り上がり、製作者が注目させる意図をもって表現されていた場面である。この場面は、多くの被験者が注目していたにもかかわらず、Level 1 になって初めて抽出されており、Link 数が少なかったことで強調されてこなかった。Link 数の算出は、オリジナル作品と複数のオマージュ作品との Cut の一致判定によって行っているため、この強調失敗はオマージュ作品選択の問題である。実験に用いた作品は、3章に分けて公開された作品であり、実験に用いたのはそのうち第 1 章だけである。対して、オマージュ作品は第 1 章のみを対象にして制作されたものもあるが、多くの作品が全章を通して表現する意図をもって制作されていると考えられる。そのため、1 章のみに最適な形で抽出はされなかった。

画像の提示枚数についての評価は、Level が下がるごとに低くなる結果を受けたが、Level 1 の時点でも 4, 3 の評価を下した人が 3 割程度おり、評価の落ち方が緩やかであった。その要因として、Level が下がるごとに増加する画像の提示枚数と各 Level で提示している範囲 (図 4) を考慮して、適当と判断されたためである。しかし、各 Level の間では不要な画像を含め提示枚数に差があった。そのため、Level が下がるごとに煩わしい印象を与え、評価を下げた。

PV との比較結果は、互いに作品全体を表現するものだが話の流れ、要所を押さえた表現といった点では PV の精度が良い、全体の把握のしやすさ、ユーザの求める場面の抽出では Level 1 が良いとされた。話の流れや要所を押さえた表現で PV が良いとされた要因は、あらかじめ要点を知り表現された PV に対して、Level 1 は複数のユーザが同様に強い印象を受けた場面を Link 数で表しているため、抽出段階でどうしても冗長な情報が

含まれやすく、不要な画像が含まれているからである。全体の把握で Level 1 が良いとされた要因は、画像の抽出範囲は Level 1 に含まれるような形で PV は抽出されており、冗長な情報は含まれるものの PV より全体を網羅する形で抽出がされていることがあげられる(図 5)。また、コメントに寄せられたように PV では物語の結の場面が不足しており、ユーザが求めている場面もまた不足していることが指摘されている。PV は、宣伝目的に公開されたものであるため、物語の結末や実際に見てもらいたい場面は伏せて物語を表現している。そのため、Level 1 よりも物語全体の表現として劣った結果となった。PV と Level 1 の抽出画像の比較を行うと、PV に含まれる画像のほぼ全部が Level 1 に含まれる画像であることが確認された。そのため、フォークソノミーによるハイライト検出で表現される Level 1 は、製作者側が演出意図をもって表現した場面を含み、物語の要所を押さえたものとなっていることが確認された。また、フォークソノミーによるハイライト検出の評価するべき点として、物語の伏線となっていた場面が抽出されたことである。その場面は、製作者側にとって、音声、エフェクトといったものを用いて特別に注目させる意図をもって表現されたものではなく、ただ物語として暗に意識させるために提示しているものである。そのような場面の抽出がされたことは物語の表現として意味を持つ。

実験によって、本手法は Level ごとに抽出範囲に見合った物語を表現していることが確認された。しかし、本手法の根幹ともいえる画像の強調の精度では高い評価を得ることはできなかった。その問題点の一つは、不要な画像の混入である。各種不要と判断される画像が、各 Level ともに 20~30%含まれるために、全体の Scene 中での集中度を偏向させ、強調を誤る傾向にあった。また、抽出に用いたオマージュ作品の影響で Link 数が十分にとられず、強調を誤る傾向も得ている。そのため、今後はそのような問題点を解決する手法を考える必要がある。

## 6. 結論

### 6.1 まとめ

本研究は、オマージュ作品に含まれるオリジナル作品へのユーザの興味情報に注目し、複数のオマージュ作品がモチーフとした場面、すなわちオリジナル作品の断片をユーザの印象に残った場面であるとする中で、ユーザの興味本意に映像メディアを要約、強調するシステムの構築、検証を行った。本手法は、そうして要約された情報をマンガコ

マ割表現にビジュアライズすることで、映像要約とユーザへの興味提示を合わせたシステムの設計を目指したものである。

本手法では、オマージュ作品とオリジナル作品の Frame の同期をとりながら、Frame ごとに多重解像度解析をかけることで Cut ごとの類似度を測り、オマージュ作品の中で使われているオリジナル作品の場面の特定を行った。その処理を複数のオマージュ作品に行うことで、オリジナル作品の中でユーザの興味が集中する場面の抽出し、その集中度をもとに映像の要約、強調を行った。その結果、映像作品の要約に対して一定の有効性がしめされた。しかし、本手法をマンガコマ割表現にビジュアライズした際に重要となる画像の強調に対しては、高い評価は得られなかった。

### 6.2 課題

本研究における実験結果、画像の強調に対して、高い評価は得られなかった。その大きな要因としては、Scene 中での集中度に偏りが出ていること、Link 数が十分にとられなかったことがあげられる。以下に課題を述べる。

#### ・ Scene に対する集中度を誤る場合

Scene 中での集中度を誤る原因は、先に述べたように主観的には不要な画像が考えられる。不要とされる画像は、2 種類に大別される。それは、背景画像、冗長と感じられる画像(節 5.2 前述)である。これらの中で、同じ画像が抽出された場合は抽出後に、類似判定をかけるなどにより改善できると考えられるが、他に關しては、ユーザの嗜好の問題も含むため、処理のみで解決を図れる問題ではない。

#### ・ Link 数が少なかった場合

Link 数が少ないことで強調を誤った場合は、抽出に用いるオマージュ作品選択の問題であるので、処理の改良で問題を解決できるものではない。

本実験によって、明らかとなった問題点は処理のみで解決を図れる問題ではない。したがって、今後は実際にシステムを完成させるに当たって、ユーザと協調して作品を完成させるシステムを設計することでこのような問題の解決を図る。また、他の素材での評価をはかり、検証を続ける。

### 6.3 今後の展望

本研究では、映像マンガ変換の前段階としてコマ画像の生成、強調に関するシステムの構築、検証を行った。従来のマンガ的に映像をマップする研究では、重要度に基づくキーフレームの強調、配置にとどまっておき、マンガの文法をとり入れることはされてこなかった。マンガは、日本で独自の発展を続けており、今では国を代表する芸術、エンタテインメントとして高い評価を受けている。それはひとえに様々な固有の表現がマンガ家同士の競争を通じて継承されてきたからである。例えば、コマ割である。マンガにとって、コマ割とは最大の特徴であり、最大の表現である。マンガ家は、ストーリーのおもたる場面を各コマに表現し、吹き出し、効果線など様々な表現をコマに盛り込むことで、一枚のイメージに様々な物語を付加している。また、表現している場面の重要度、読者の思考を考慮した興味を引くコマ配置に工夫を凝らしている。このような多彩な表現を取り込むことで、より効果的な映像ダイジェストが可能になるだけでなく、一つのエンタテインメントとして、ユーザの興味を引き付ける表現が可能になる。今後、研究として取り組む課題を以下にあげる。

#### ・tag 情報を加味した抽出

本研究の実験では、抽出素材の選択に作品タイトルで検索され、かつ再生数の多いものを選択した。しかし、同じ作品を用いたオマージュ作品でも様々な tag 付けをされており、その性質にそくした鑑賞者の評価が tag に見ることができる。そのため、抽出に用いる素材の選択に tag の情報を加えることで、どのように抽出される画像が変化するか検証する必要がある。

#### ・画像のトリミング

オリジナル作品より画像を取り出した場合、すべての画像はオリジナル作品の規格に合わせたサイズである。マンガの「コマ」に収めるためには、抽出された画像のトリミングが必要となる。

#### ・マンガの文法に即した変換

映像作品とマンガ表現は、表現形態が異なる。そのため、マンガ的表現に変換するために、抽出された画像を選別すると共に、紙媒体への印刷を前提として発展してきたマンガ表現の「文法」を採り入れた配置、演出が求められる。

### 一見開き単位の場面構成とコマ割のリズム

マンガは「めくる」といった紙媒体ならではの行為を前提とした演出が行われている。そのため、見開きの1ページを意識して場面構成を成立させる必要があり、同時に次の1ページを「めくる」楽しさを与えるものでなければならない。例えば、はじめと最後のコマは読者をひきつけるものを配置することである。また、読者をひきつけるコマにつなげる見せ場を見開きの1ページに成立させる必要がある。そうした演出に、重要な場面をコマのサイズを変えるだけでなく、緩急ある物語を感じさせるコマの切り出しも求められる。

### 一効果線、吹き出しといったマンガ表現

マンガらしい表現を突き詰めるなら、コマに選択されたイメージにあわせた効果表現をのせることも考える必要がある。また、本研究ではテキスト情報を扱うことはしていないが、吹き出しといった表現も考えるなら、イメージの持つテキスト情報を取得する方法が必要である。

### 参考文献

- [1]内橋真吾, “ビデオ・マンガ要約を用いたインタラクティブなビデオ閲覧”, インタラクション 2001 論文集, pp31-32, 2001.
- [2]中村貴洋, 青木秀憲, 宮下芳明, “マンガ的手法を用いたニコニコ動画ナビゲーション”, ヒューマンインタフェース学会研究報告書 Vol.10, No.4, 2008.
- [3]山本大介, 長尾確, “半自動ビデオアノテーションとそれに基づく意味的ビデオ検索”, 情報処理学会第 65 回全国大会, 2003.
- [4]椋木雅之, 森田裕士, 馬場雅志, 浅田尚紀, “シーン構成グラフを用いたドラマ映像の構造表現”, 電子情報通信学会技術報告, MVE, Vol.106, No.234, pp.43-48, 2006.