

文脈を考慮に入れた映画の要約作成

加藤 和也 吉高 淳夫 平川 正人

広島大学大学院工学研究科
〒739-8527 東広島市鏡山1丁目4番1号
{kazya, yoshi, hirakawa}@isl.hiroshima-u.ac.jp

本稿では、音と画像の特徴を利用し、映画の文脈に注目した要約作成手法を提案する。一般に、人が映画を見る時、その映像区間だけでなく前後のつながりから内容を理解している。そのため、映画を要約するためには文脈を考慮に入れるべきである。しかし、従来の映画を対象とした映像要約の研究は、特定の特徴を示す区間を単純に検出したものがほとんどで、文脈を考慮に入れたものは報告されていない。本研究では、映画の文法に基づいて、音、画像の特徴が類似しているショットの連続をストーリー・ユニットとして検出し、ショットの従属性、ストーリー・ユニットの従属性を映画の文脈として検出する。文脈を考慮に入れることで、ストーリーの流れを損なわずに映画を要約することができる。

Video Summarization based on Movie Context

Kazuya Kato Atsuo Yoshitaka Masahito Hirakawa

Graduate School of Engineering, Hiroshima University
4-1, Kagamiyama 1 chome, Higashi-Hiroshima, 739-8527
{kazya, yoshi, hirakawa}@isl.hiroshima-u.ac.jp

In order to summarize a movie, movie's context should be taken into account. However, past research focused only on detecting sections that show specific features, and relationships (i.e., context) among those sections are not taken into account. In this paper, a framework of video summarization based on movie's context is proposed. A movie is segmented into story units each of which is the sequence of shots with similar audio and visual features. We regard movie's context is dependency between shots / story units. Considering the shot pair and story unit pair with strong dependency, a movie is summarized without spoiling the flow of story.

1. はじめに

コンピュータ上で映像メディアを利用する機会が増加しており、インターネットによる映像配信やデジタル放送が一般化しつつある。その結果、一般視聴者にとって、膨大な量の映像の中から見たい映像を選択する必要性が高まっている。そのため、映像の内容を短時間で

把握する手法として、要約映像を自動作成する研究が、これまでにいくつか報告されている[1][2][3][4]。映画を対象とした研究として文献[3]では、特別なイベント(主要人物のクローズアップ、銃声や爆発、タイトルやテロップなど)を検出し、これらをつなぎ合わせることで映画の予告編を目的とした要約映像を作成している。また

文献[4]では、ドラマの心理的印象の高い区間に注目し、音楽の開始や終了、カットが頻出する箇所などを検出することで心理的に重要な箇所のみを切り出した要約映像を作成している。これらの手法で検出されている要約映像は、その時間毎に重要な区間を検出し、それを単純につなぎ合わせているに過ぎない。一般に人が映画を見るとき、その映像区間だけでなく、前後の映像区間との関連に基づいて内容を理解している。そのため、映画の内容を理解しやすい要約映像を作成するためには、文脈を考慮に入れるべきである。しかしながら、映画を対象とし、その文脈を考慮に入れた要約作成に関する研究はこれまでに報告されていない。

本稿では、対象とする映像を映画に限定し、そこに含まれる音と画像、および映画の編集技法[8]上の特徴から映画の文脈を分析し、この結果から要約を作成する手法を提案する。

2. 映画の文脈

2.1 映画の構造化

映画から文脈を抽出し要約映像を作成するために、まず映画を構造化する。そこで、文献[5]の手法を利用して映画をショット、シーンに分割する。ショット、シーンは以下のように定義する。

- ・ショット：単一のカメラから撮影された連続するフレームの集合
- ・シーン：同じ場面(撮影場所)での出来事を表した連続するショットの集合[5]

本研究では、映画の文脈に基づいて、画像、音の特徴が類似しているショットの連続をストーリー・ユニットとしてシーン内から検出する。映画の構造を図1に示す。

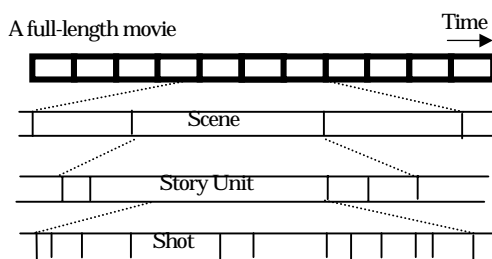


図1 映画の構造

2.2 映画の文法

映画には、撮影や編集の際に制作者によって特定の意味や意図を強調する目的で使用される技法がある[8]。本研究では、これらの映画の編集、撮影時の特徴に基づいて要約映像の作成を行う。そこで、各ショットから抽出する特徴をショットの長さ、ショット内の画像変化、及び類似ショット対とする。ショット内の画像変化は、カメラやオブジェクトの動きを表し、類似ショット対は、ショット内のフレームが類似しているショット対とする。

それぞれの特徴について、以下に示す

(1) ショットの長さ

短いショットが連続することで、緊迫した場面を表す。また、徐々にショットの長さが短くなる時緊迫感の高まりを、逆に徐々にショットの長さが長くなる時緊迫感からの解放を表す。

(2) ショット内の画像変化

画像変化が大きいショットが連続するとき、カメラやオブジェクトによる動きが連続している可能性が高い。

(3) 類似ショット対

映画では、同一シーン内で類似したショットが繰り返し出現する機会が多い。特に、登場人物同士の対話区間では、登場人物のアップ等、類似したショットが繰り返される。

2.3 本研究における文脈の定義

一般に、文脈とは文章の流れの中にある意味的なつながりぐあいを意味する。しかし映画から、そのストーリー展開など、意味的な内容を自動的に抽出することは難しい。しかし、映画の文法を考慮に入れることで、映画から大まかな意味内容を抽出することができる[7]。そのため、映画の文法に基づいて、画像、音の特徴から個々のショット、個々のストーリー・ユニットの関係を検出することが可能である。そこで、ショットとショットの間、ストーリー・ユニットとストーリー・ユニットの間に強いつながりがあるとき、そのショット対は従属関係にあり、強い従属関係にあるショット対、ストーリー・ユニット対は互いに不可分な関係にあるとする。本研究では、映画の文脈を個々のショット及び個々のストーリー・ユニットの間の従属関係と定義する。

3. 処理概要

処理手順の概略を以下に示す。本研究では、入力した要約時間長に近い長さをもつ要約映像を作成する。

- (1) ショットの従属性を検出する。
- (2) ショットの従属性に基づいて映画をストーリー・ユニットに分割する。
- (3) ストーリー・ユニットの重要度を算出する。
- (4) 重要度の高いものから順にストーリー・ユニットを選択し、(4a)、(4b)の処理によって、要約映像として採用する区間を決定する。
 - (4a) ストーリー・ユニットから、ショットの従属性に基づいて要約映像に採用するショットを検出する。
 - (4b) 選択したストーリー・ユニットに対して、従属関係にあるストーリー・ユニットが存在するとき、同様にそのストーリー・ユニットからも要約映像に採用するショットを検出する。
- (5) 要約映像として選択した映像の合計時間長が目的の長さを越えるまで(4)の処理を繰り返す。
- (6) 選択された主要なショットを時間順に並べ、要

約映像とする。

4. 映像の特徴抽出

4.1 ショット内の画像変化量

各ショットの画像変化量を以下のように定義する。
 $Y(x, y, i)$ ($0 \leq Y(x, y, i) \leq 1$) は、フレーム i 内の画素 (x, y) の輝度値とする。

$$\Delta Y(x, y, i) = |Y(x, y, i+1) - Y(x, y, i)|$$

$$D(i) = \frac{1}{N} \sum_{x=1}^{l_x} \sum_{y=1}^{l_y} 1, \text{ if } \Delta Y(x, y, i) > \Delta Y_{th}$$

l_x, l_y はフレームの x 軸、 y 軸方向の画素数、 $N = l_x \times l_y$ である。フレーム $i, i+1$ 間の輝度値の差 $\Delta Y(x, y, i)$ に対する閾値は $\Delta Y_{th} = 0.05$ とする。

そして、ショット s_k の画像変化量 $SA(s_k)$ を以下の式で算出する。 h, b はそれぞれショット s_k 内の最初と最後のフレーム番号を表す。

$$SA(s_k) = \frac{1}{b-h} \sum_{i=h}^{b-1} D(i)$$

4.2 類似ショット対の検出

4.2.1 キーフレーム間の類似度の算出

類似ショット対の検出には、キーフレーム間の類似度を用いる。以下のフレームをキーフレームとする。

- ・ ショットの先頭フレーム
- ・ ショットの最終フレーム
- ・ ショットの先頭から 1 秒毎のフレーム

キーフレーム間の類似度を以下の様に算出する。まず、比較するキーフレーム同士を x 軸方向、 y 軸方向共に 8 等分し、分割後の領域毎にカラーヒストグラムを比較する。まず、YIQ 色空間 ($0 \leq Y \leq 1, -0.6 \leq I \leq 0.6, 0.52 \leq Q \leq 0.52$) を人間の視覚特徴に応じて 16, 6, 2 等分した、全体で $16 \times 6 \times 2$ の 192 個の部分空間に分割する。そして、キーフレーム k_1 内の領域 $R(rx, ry)$ ($rx = 1, \dots, 8, ry = 1, \dots, 8$) に対して、 i 番目の部分空間 $c_i (i=1, \dots, 192)$ に含まれる画素数を $h_{k_1 R(rx, ry)}(c_i)$ とする。

キーフレーム k_1, k_2 間の類似度 $V(k_1, k_2)$ を以下の式で算出する。

$$V(k_1, k_2) = \frac{1}{64} \sum_{rx=1}^8 \sum_{ry=1}^8 \left(1 - \frac{1}{8 \times \frac{8}{8}} \sum_{i=1}^{192} \left| h_{k_1 R(rx, ry)}(c_i) - h_{k_2 R(rx, ry)}(c_i) \right| \right)$$

4.2.2 ショット間の類似度

キーフレーム間の類似度から、ショット間の類似度を算出する。ショット s_k, s_m 内のキーフレーム数をそれぞれ n_{s_k}, n_{s_m} とし、ショット s_k の i 番目のキーフレームを $f_{s_k i}$ 、 s_m 内の j 番目のキーフレームを $f_{s_m j}$ とする。

このとき、以下の式でショット間の類似度 $KS(s_k, s_m)$ を算出する。

$$KS(s_k, s_m) = \frac{1}{n_{s_k} \times n_{s_m}} \sum_{i=1}^{n_{s_k}} \sum_{j=1}^{n_{s_m}} 1, \text{ if } V(f_{s_k i}, f_{s_m j}) > V_{th}$$

$V(f_{s_k i}, f_{s_m j}) > V_{th}$ のとき、キーフレーム対 $f_{s_k i}, f_{s_m j}$ が類似しているとして、ショット間の類似度 $KS(s_k, s_m)$ 類似キーフレーム対の数とする。このとき、 $V_{th}=0.5$ とした。

4.2.3 類似ショットによるラベリング

同一シーン内の類似ショットを以下のアルゴリズムでラベリングする。

```
for k=1 to num{
  if((SA(s_k) > SA_th) or (k=1))
    label[s_k] 新規ラベル
  else{
    label[s_k]  null
    for i=1 to k-1{
      if((KS(s_k, s_{k-i}) > KS_th) and (SA(s_{k-i}) ≤ SA_th)){
        if(KS(s_k, s_{k-i}) > KS_th)
          label[s_k] label[s_{k-i}]
      }
    }
    if(label[s_k] ≠ null)
      break
  }
  if(label[s_k]=null)
    label[s_k] 新規ラベル
}
```

num: シーン内のショット数

$KS(s_k, s_j)$: ショット s_k, s_j の類似度

$SA(s_k)$: ショット s_k の画像変化量

KS_{th} : ショット間の類似度 $KS(s_k, s_j)$ に対する閾値

SA_{th} : 画像変化量 $SA(s_k)$ に対する閾値

s_k : ショット s_{k-i} と同一のラベルを持つすべてのショット

label[s_k]: ショット s_k のラベル

閾値はそれぞれ、 $SA_{th}=0.2, KS_{th}=0.25$ とした。このアルゴリズムにより、同一シーン内のすべての類似ショットに対して、同一のラベルを付加することができ、同一ラベルを持つすべてのショットの組み合わせに対して、 $KS(s_k, s_j) > SA_{th}$ である。

4.3 音の特徴抽出

文献[6]の手法により、音楽、効果音(SE)、音声が入入されている区間をそれぞれ検出する。

SE に関しては、映画には数多くの SE が挿入されているため、本研究では音量が一定値以上の SE のみに注目する。SE の音量 $Loud$ は以下の式で算出する。

$$Loud = \max \left(\frac{1}{S} \sum_{i=0}^S |A_i|, \frac{1}{S} \sum_{i=S}^{2S} |A_i|, \dots, \frac{1}{S} \sum_{i=N-S}^N |A_i| \right)$$

N : SE 区間のサンプル数

SE 区間内の音波形について、 i 番目のサンプルの振幅を A_i とし、一定サンプル数 S 毎に振幅の絶対値の平均を求める。その最大値を SE の音量 $Loud$ とする。

5. ショットの従属性の検出

5.1 画像の特徴によるショットの従属性

ショット内の画像変化量 $SA(s_k)$ 、ショットの長さ $L(s_k)$ [秒] から、従属関係にあるショット間を検出する。以下の条件を満たす連続するショット s_k, \dots, s_{k+n} ($n \geq 2$) では、ショット s_{k+i} ($i=1, \dots, n$) は直前のショット s_{k+i-1} に対して従属しているとする。

- (i) $SA(s_k), \dots, SA(s_{k+n}) \geq 0.1$ and $L(s_k), \dots, L(s_{k+n}) \leq 4$
- (ii) $L(s_k), \dots, L(s_{k+n}) < 2$ or $L(s_k) > L(s_{k+1}) > \dots > L(s_{k+n})$
- (iii) $L(s_k) < L(s_{k+1}) < \dots < L(s_{k+n})$

また、以下の類似ショット対は従属関係にあるとする。

- (iv) ショット s_k と同一ラベルのショットが、ショット s_k 以前の 3 ショット以内、 s_{k-3} までに存在するとき、ショット s_k は同一ラベルをもつ直前のショットに対して従属しているとする。ただし、ショット s_k から s_{k-3} までの時間長の合計が 20 秒に満たない場合は、20 秒を満たすショットまで調べる。

5.2 音の特徴によるショットの従属性

音の特徴に注目し、(v) 音声、(vi) 効果音(SE)、(vii) 音楽から、従属関係にあるショット対を検出する。

(v) 音声による従属性

文献[9]で、発話の際の言葉の「間」は、比較的長い「間」で約 1.4 秒であることが報告されている。そのため、本研究では音声と音声の「間」が 1.4 秒以内のとき、その 2 つの音声は連続していると判断する。カットを挟んで音声は連続しているとき、その連続するショットは従属関係にあるとする。

(vi) SE による従属性

カットの直前で音量が閾値以上の SE が挿入されているとき、その次のショットは SE に対するリアクションを表していると考え、その連続するショットは従属関係にあるとする。

(vii) 音楽による従属性

映画における音楽は、その場面の雰囲気や登場人物の心境を強調するために映画編集者によって挿入されたものである。そのため、同一の音楽区間内のショットでは、同じような雰囲気または登場人物の心境が描かれていると考えられる。そのため、ショット s_k からショット s_{k+n} にまたがって音楽が連続しているとき、ショット s_{k+i} ($i=1, \dots, n$) はその直前のシ

ョット s_{k+i-1} に対して従属関係にあるとする。

6. ストーリ・ユニットの検出

ショットの従属性に基づいて、ストーリー・ユニットを検出し、シーンをストーリー・ユニットに分割する。以下の順に処理を行う。

- (1) まず、緊迫感の高いショットの連続をストーリー・ユニットとして検出する。そこで、5.1 の(i), (ii) によって従属関係にあるショットの連続を検出する。このとき、(i), (ii) の区間が重複して検出される場合はそれぞれの区間を結合し、ひとつのストーリー・ユニットとする。
- (2) 次に(1)に含まれないショットから類似ショットの繰り返しを検出する。そこで、5.1 の(iv)より、類似ショットとして従属関係にあるショット対 s_k, s_{j+i} を検出する。ショット s_{j+i+1} 以降のショットについても同様に、従属関係にあるショットを検出する。このとき、これらの従属関係にあるショット対とその間に挟まれているショットを結合することで類似区間を検出する。

そして類似区間に対して、5.1 の(iii)、5.2 の(vii) により従属関係にあるショットが存在するとき、それらを結合することで、ひとつのストーリー・ユニットとする。

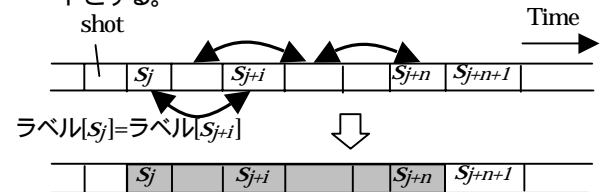


図2 類似区間の検出

- (3) (1)(2)に含まれないショットからは音楽に注目し、同一の音楽区間内のショットをひとつのストーリー・ユニットとして検出する。(1)(2)によって検出したストーリー・ユニットの境界が、音楽区間に存在するときは、そのストーリー・ユニットの境界を優先する。ショット s_m が音楽の開始を含むとき s_{m-1}, s_m 間のカットを、ショット s_m が音楽の終了を含むとき s_m, s_{m+1} 間のカットを、それぞれストーリー・ユニットの境界として検出する。検出したストーリー・ユニットの境界によって、シーンをストーリー・ユニットに分割する。

7. 要約映像の作成

7.1 ストーリ・ユニットの重要度の算出

ストーリー・ユニットに対して重要度を算出する。ストーリー・ユニットの重要度は、前後のストーリー・ユニットと比較して、その特徴がより大きく異なるストーリー・ユニットほど重要とすることで算出する。ここでは、ストーリー・ユニット u_m の特徴 $LA(u_m)$ を u_m 内の各ショットの平均時間[秒]とした。ただし、 $LA(u_m) > 4$ のとき、

$LA(u_m)=4$ とする。そして、ストーリー・ユニット u_m の重要度 $I(u_m)$ を以下の式で算出する。

$$I(u_m) = |LA(u_m) - LA(u_{m-1})| + |LA(u_{m+1}) - LA(u_m)|$$

7.2 ストーリー・ユニットの従属関係

まず、ストーリー・ユニットを以下の(A)(B)に分類する。

- (A) 6の(1)によって検出した、緊迫感の高いストーリー・ユニット
- (B) その他のストーリー・ユニット

7.1 で選択したストーリー・ユニットに対して、直前、直後のストーリー・ユニットが以下のどちらかの条件を満たすとき、そのストーリー・ユニット対は従属関係にあるとし、従属関係にあるストーリー・ユニットも同様に要約映像に選択する。

- ・ 連続するストーリー・ユニットが、同一の音楽区間内に含まれる。
- ・ 直前、または直後のストーリー・ユニットの特徴が異なる。

7.3 要約映像として採用する映像の検出

7.1, 7.2 で選択したストーリー・ユニットを代表する数ショットを検出し、そのショットを要約映像に採用する。そのために、まずストーリー・ユニットの代表ショットを1つ検出し、代表ショットと従属関係にあるショットを連結していくことで、要約映像に採用するショットを決定する。ストーリー・ユニット内の代表ショットは以下の順に決定する。

- (1) ストーリー・ユニット内で音楽の開始、または終了を含むとき、そのショットを代表ショットとする。
- (2) (1)に含まれないストーリー・ユニットのうち、7.2の(A)に分類されるストーリー・ユニットは、SE から代表ショットを決定する。SE 以前の5秒間の音量の平均を算出し、SE の音量との差を求める。この値が最大となるSE を検出し、そのSE を含むショットを代表ショットとする。
- (3) (1)(2)に含まれないストーリー・ユニットに関して、ストーリー・ユニットの先頭から2秒毎のフレームに対して、そこから5秒間に含まれる音声の存在している時間を求める。ストーリー・ユニット内で、その値が最大となるフレームを求め、そのフレームを含むショットを、ストーリー・ユニットの代表ショットとする。

代表ショットと従属関係にあるショットを連結し、要約映像として採用するショットを決定する。このとき、従属関係にあるショットが存在しなくなるまで、もしくはストーリー・ユニット u_m から選択したショットの合計時間が、時間制約として設定した T [秒] を超えるまでその処理を繰り返す。ここでは、 $T=10$ とした。

8. 実験結果

2本の映画に対して、約5分、約10分、約20分の要約を提案手法によって作成し、実験を行った。なお、使用したビデオデータの形式は、フレームサイズ 160×120 [pixel]、フレームレート 30 [frames/sec.]、24ビットカラー、オーディオ形式はサンプリング周波数 22.050 [kHz]、量子化8ビット、モノラルである。

8.1 比較対象とする映像要約手法

比較対象として、以下のようなカットの頻度による要約映像を作成した。映画の先頭から5秒毎のフレームに対して、そこから10秒間に含まれるカットの数を求める。この10秒間に含まれるカット数が最も多いフレームから順にキーフレームとする。キーフレームが含まれるショットを先頭ショットとして、先頭ショットから合計時間が10秒を越えるまでのショットを連結し、要約映像として採用する。要約映像の時間長が目的の時間に達するまでその処理を繰り返し、選択した区間を時間順に並べることで要約とした。

8.2 評価方法

提案手法による要約映像を評価するため、大学生3名の被験者に、映画から重要と思うショットを列挙させることで、それぞれの被験者毎に20分の要約映像を作成した。このときすべての被験者が重要と判断したショットに注目する。そのショットは、他のショットと比較して、映画において重要である可能性が高いと考えられる。そこで、3名の被験者全員が重要と判断したショットに対して、要約映像に含むことのできたショットの割合を R_c として以下のように評価する。

$$R_c = \frac{C_s}{A_s}$$

A_s : 被験者3名全員が重要と判断したショット数
 C_s : A_s のうち要約に含むことのできたショット数

次に、文脈を考慮に入れることで、作成した要約が不自然な箇所では区切れていないか以下のように評価した。そこで、要約映像として切り出した映像区間の開始、終了のカットについて、そのカットの前後のショットが従属関係にあるか、実際にその前後の映像を見ることで判別した。カット前後のショットが、従属関係にあると判別された場合、そのカットは要約映像として不適切である。そこで、 R_f として以下のように評価する。

$$R_f = \frac{F_b}{T_b}$$

T_b : 要約映像として、切り出した映像区間の開始、終了のカット数
 F_b : T_b のうち、前後のショットが従属関係にあると判別されたカット数

R_f は0%に近いほど良い要約であり、100%に近い高いほど、中途半端な箇所では区切れているため、その要約は内容が分かりにくい要約といえる。

実験結果を以下の表 1, 2 に示す。

表 1 評価に利用した映画サンプルのデータ

	ショット数	シーン数	ストーリー・ユニット数	時間[min]
Sample1	1420	40	214	104
Sample2	1481	49	210	118

表 2 実験結果

	<i>Rc</i>			<i>Rf</i>		
	5分	10分	20分	5分	10分	20分
Sample 1	21.6%	28.9%	39.2%	21%	29.7%	31.2%
	21/97	28/97	38/97	8/38	19/64	86/125
Sample 2	1.1%	21.4%	39.2%	53%	39.3%	41.7%
	1/84	18/84	33/84	18/34	26/68	57/137

(a)提案手法による要約

	<i>Rc</i>			<i>Rf</i>		
	5分	10分	20分	5分	10分	20分
Sample 1	20%	28.9%	36%	21.9%	22.3%	22.3%
	19/97	28/97	35/97	7/32	12/54	20/90
Sample 2	19%	35.7%	54.7%	37.5%	47.2%	43.9%
	16/84	30/84	46/84	20/32	33/70	43/98

(b)カットの頻度による要約

8.3 考察

提案手法と比較対象として用意したカットの頻度による要約手法との間で、結果に大きな違いは見られなかった。提案手法ではストーリー・ユニットの重要度をショットの長さのみで算出しているため、カットの頻度が高いストーリー・ユニットが多く選択された。そのため、両方の手法で共通して選択されたショットが数多く存在し、特に、要約の時間を長く設定した20分の要約映像では、それぞれの手法で検出したショットのうち、提案手法によって検出したショットの半数近くのショットが、比較対象の要約と共通して選択されていた。そのため、*Rc*に関してほぼ同じような結果となった。ただし、*Rc*の結果については、3名の被験者全員が重要だと考えたショットが10分の要約に20%以上、20分の要約には40%以上含まれている。このことは、元の映画データが共に100分以上であることから考えると、高い数字だと考えられる。

*Rf*に関しては、提案手法は良い結果が得られていない。比較対象として用意したカットの頻度による要約は、文脈に関して全く考慮に入れていないのにもかかわらず、提案手法は、この手法と同程度の結果しか得られていない。そのため、提案手法による要約映像は、この評価方法において、映画の文脈を考慮に入れることができているとはいえない。

9. まとめ

本稿では、画像、音の特徴から、映画を構造化し、それぞれの単位における従属性を検出することで、映画の

文脈を考慮に入れた映像要約手法を提案した。そして、提案手法について、文脈を考慮に入れていない要約映像を作成し、比較することで評価を行った。

今後の課題として、まず評価方法について検討する必要がある。また、要約作成手法について、ストーリー・ユニットの選択手法、ストーリー・ユニット内からのショットの選択手法について検討し直す必要がある。

謝辞

本研究を行うにあたり、多くの意見、助言、多大な協力を頂いた出口嘉紀氏に心から感謝します。

参考文献

- [1] Michael A. Smith, Takeo Kanade, "Video Skimming and Characterization through the Combination of Image and Language Understanding", IEEE International Workshop on Content-based Access of Image and Video Databases, pp. 61-70, 1998.
- [2] Alan Hanjalic, HongJiang Zhang, "An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 9, No. 8, December 1999.
- [3] R. Lienhart, S. Pfeiffer, W. Effelsberg, "Video Abstracting", Communications of the ACM, Vol. 40, No. 12, pp. 55-62, Dec. 1997.
- [4] 森山剛, 坂内正夫, "ドラマ映像の心理的内容に基づいた要約映像の生成", 電子情報通信学会論文誌, Vol. J84-D-II, No. 6, pp. 1122-1131, Jun. 2001.
- [5] Yoshitaka, M. Miyake, "Scene Detection by Audio-Visual Feature", Proc. of IEEE International Conference on Multimedia and Expo., pp. 49-52, 2001.
- [6] 川崎智広, "A Method of Extracting Sound Components in Films and Evaluating their Impressions", 広島大学修士論文, 1998.
- [7] Yoshitaka, T. Ishii, M. Hirakawa, and T. Ichikawa, "Content-Based Retrieval of Video Data by the Grammar of Film," Proc., IEEE Symposium on Visual Languages (September 1997), pp. 310-317.
- [8] ダニエル・アリホン著, 岩本憲児, 出口丈人訳, "映画の文法", 紀伊國屋書店, 1980.
- [9] 中村敏枝, "間(ま)の感性に関する心理学的研究", 信学技報, HC92-71, March 1993.