5ZD-7

音楽理論に基づく映画の要約映像生成手法

竹内 星子† 浜中 雅俊‡

筑波大学システム情報工学研究科† 筑波大学システム情報系‡

1. はじめに

本稿では,音楽理論 Generative Theory of Tonal Music (GTTM)[1]に基づき映画を構造化し,要約映像を生成する手法を提案する.映画の要約の研究では,内容の充実度と制約時間のトレードオフと,個人によって生じる要約の目的の違いに着目し,以下の三つの要件のいずれかに焦点を当てているものが多かった[2-4].

1) 映画の内容が理解できる
2) 要約映像をユーザの求める時間長に収める
3) ユーザの関心の強い場面をまとめる

文献[2,3]では,上記の要件の 1 と 2 を満たすことを目的としていたが,重要な場面をつなぎ合わせるだけで,各場面同士に因果関係がなく,要件 1 を十分に満たしてはいなかった.一方,要件 1 に焦点を当てた研究[4]は,ユーザが各場面の内容と連続する場面同士の因果関係を記述する方法を提案した.すべての要件に焦点を当てた研究[5]は,音声箇所と非音声箇所の再生速度を変え,視聴時間を短縮する方法を提案した.しかし,[4,5]の方法では,映画の内容を正確に把握したり,高速な映像を視聴したりするためのユーザの負担が大きかった.

そこで我々は,GTTM による楽曲の簡約を応用し,同じように時系列メディアである映画を要約する方法を提案する.楽曲の簡約には,重要な音と装飾的な音の従属関係を表すタイムスパン木を用いるが,本研究では映画の大局的構造と局所的構造それぞれのタイムスパン木を獲得し,要件のすべてを満たす要約が可能か検証する.

実験の結果,上記の要件の 1 と 2 を満たすような要約映像を生成できることを確認した.

2. 映画の構造化

本研究では,音楽理論 GTTM を応用した構造化によって映画の重要な場面と各場面の従属関係を表すタイムスパン木を獲得し,大局的構造と局所的構造に分けることで,第 1 章で述べた要件すべてを満たす映画の要約を目指す.

"Study of video summary generation method with structured movie" †Takeuchi Seiko, University of Tsukuba, Graduate School of Systems and Information Engineering, ‡Hamanaka Masatoshi, University of Tsukuba, Faculty of Engineering, Infomation and Systems.

2.1. 大局的・局所的構造のタイムスパン木

映画では,時系列的に離れた場面同士に重要な因果関係が存在することがある.人間が内容を理解するためには,意味的な内容を持つ映像が必要である.そのため,数秒で構成されるような短い映像のみの従属関係で作成されたタイムスパン木では内容を理解するのは困難である.そこで我々は,大局的構造,局所的構造を表すタイムスパン木を獲得する方法を提案する.

映画において映像の最小単位の映像とされているショットでは局所的構造を表すタイムスパン木を,関連する連続したショットであるシーンでは,大局的構造を表すタイムスパン木を生成する.シーンのタイムスパン木によって意味的な内容を持つレベルで要約を行い,ショットのタイムスパン木では,それだけでは意味を持たないような映像を用いた細かな調節を行う.

2.2. 構造化の手順

音楽理論 GTTM は音楽の三要素である旋律,リズム,和声を用いて楽曲の分析を行い,図 1a のようなタイムスパン木を獲得する.我々は,楽曲では音符からフレーズを,映画ではショットからシーンという様に,ゲシュタルトを形成するという類似点に着目し,以下の手順で構造化を行い図 1b のようなタイムスパン木を獲得する.

1) ショットをシーンにグルーピング
   新たなシーンが開始すると考えられる場面に共通点を見つける.たとえば,繰り返し類似するショットが出現する場合や,人物が一変したりする場面などがある.
2) シーンのタイムスパン木を獲得
   ストーリが進行,あらすじを説明する情報量の多い場面などを重要とし幹を生成,前のシーンの補足のみ,内容が重複する,かつ情報量が少ない場面などを重要でない場面とし,幹に従属する枝を形成する.
3) ショットのタイムスパン木を獲得
   場所の変化,人物が入退場する場面などを重要とし幹を形成,反応のみの場面などを重要でない場面とし幹に従属する枝を形成する.

どのような場面を重要とするかによってタイムスパン木の形は変化する.たとえば、特定の人

物に重点を置きたい場合には,その人物が出てくる木が優先的に幹に選択されるようにする.したがって,ユーザの関心の強い場面をまとめるという第1章の要約の要件3を満たすことができる.

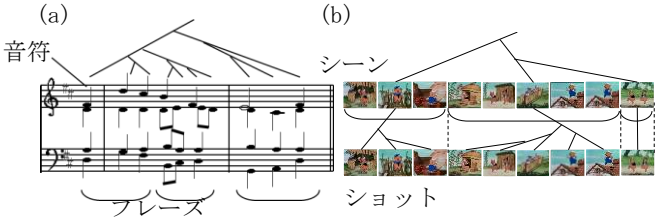

図 1.楽曲と映画のタイムスパン木

3. 映画の要約

我々は,楽曲の簡約をタイムスパン木の深さによって決定する方法を映画に応用した.ここで,楽曲の簡約は,音同士の重要度を比較し,重要な音を選択して段階的に音の数を減らす作業を指す.これを応用し,映画の要約では,制約時間長に映像を収めるために,重要でない内容を段階的に削っていく作業を実現する.

設定した深さとの交点より深い枝を省略し,楽曲の簡約では図2a,映画の要約では図2bを得る.第1章の要件1を満たすには,シーンのタイムスパン木を獲得,操作することで可能である.また要件2,3を満たすには,ショットのタイムスパン木を編集する必要があるが,この場合においても,シーンのタイムスパン木が大局的構造を保つため,要件1に大幅な影響を及ぼすことはない.

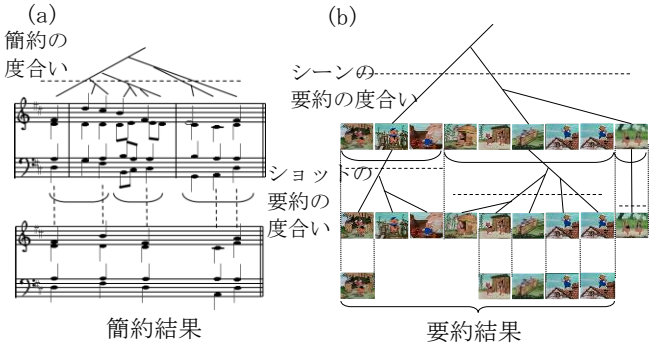

図 2.楽曲の簡約と映画の要約

4. 実験結果

映画のタイムスパン木から内容が理解できる要約が可能であるか実験によって検証した.まず,約 8 分の短編映画「三匹の子ぶた」(ウォルト・ディズニー,1933 年)から,タイムスパン木の深さ別に,1 分から 8 分まで 1 分おきの要約映像を 8 種類生成した.そして 20 代の男性 4 人と女性 1 人が,8 種類の要約映像を短い順に視聴し、一つの要約映像を視聴するごとに以下の 3 点を意識した20の設問に回答した.

- 各場面の因果関係を理解しているか
- 人物の心情を読み取ることができているか
- 映画全体の内容を理解することができたか

例えば,設問の一つである「子ぶたが恐れている動物はなぜ子ぶたを追いかけてきたか?」は、台詞での説明がないが、前後の文脈から推測することで解答できるようになっている.

その結果,要約映像が長くなるにつれて得点も増加した.また,2 分以上の要約映像に対して,半分以上設問に回答できていた.また,元の 8 分の映像を半分の 4 分に要約した結果,得点は 71 点となり,7 割以上を理解できていた. したがって,本手法による要約映像は,時間に対する内容の理解度が高いといえる.

表 1.設問に対する平均得点(100点満点)

時間(分)	1	2	3	4	5	6	7	8
平均得点	33	56	63	71	72	79	90	93

5. まとめ

音楽理論 GTTM を応用して映画を構造化し,重要度や従属関係を表すタイムスパン木から要約映像を作成する手法を提案した.また,大局的構造と局所的構造を持つタイムスパン木を作成することで,第 1 章で述べ要件すべてを満たす映画の要約が可能か検証した.

実験の結果,映画の構造化し,タイムスパン木によって要約する方法は,要件を満たすために有効であることが確認できた.

参考文献

[1] Lerdah,F.and Jackendoff,R: “ A Generative Theory of Tonal Music “, the MIT Press, Cambredge,1983.

[2] オン・コックメン,大野 雄也,亀山 渉, “瞳孔径・視線と心拍情報を用いた映像要約方法とその評価”,電子情報通信学会論文誌 A, vol.J93-A, NO.11, pp.697-707, 2010.

[3] 出口 嘉紀,吉高 淳夫, “映画の文法に基づく要約映像の生成”データベース・システム研究報告 DBS-132,pp33-40, 2004.

[4] 堀内 直明,上原 邦明, “ストーリの内容記述に基づく映像の検索と要約”,電子情報通信学会技術研究報告.DE, データ工学 97(161), pp73-78,1997.

[5] 栗原 一貴,佐々木 洋子,緒方 淳,後藤 真孝, “音声区間自動検出技術を用いた変則再生方式による映像の高速鑑賞システムの検討”,情報処理学会研究報告 Vol.2012-HCI-149, No.13,2012.