2007－HCI－124 (12)
2007／7／7

映画等を対象としたダイジェスト映像生成のための映像特徴に関する検討

吉高淳夫，田中壮詩，平嶋宗

広島大学大学院工学研究科

放送の多チャンネル化や大容量HDDを搭載したビデオレコーダの普及などにより，アクセス可能な映像コンテンツの量は増加の一途をたどっている．このような状況において，映像コンテンツの選択を支援するためのダイジェスト映像の自動生成は映像へのアクセス性を高めるための重要な技術の１つと考えられる．本稿では，映画やドラマなどの映像コンテンツを対象とし，制作者がある場面を演出するために用いる技法とその技法を適用した結果，映像，音声に現れる特徴に着目する．そして，映像に付与される撮影・編集上の演出が視聴者への印象付けにどのように寄与するかに関する実験を行い，効果的なダイジェスト生成へ向けた指針について考察した．

Investigation on Audio-Visual Features for Creating Movie Digest

Atsuo Yoshitaka, Takeshi Tanaka, and Tsukasa Hirashima

Graduate School of Engineering, Hiroshima University

The amount of video contents that we can access is growing because of the multichannel broadcasts and widespread use of HDD mounted video recorders. Video digesting for end users' contents selection is considered to be one of the important technologies for improving accessibility of video contents under this circumstance. In this paper, we focus on the directing on movie or drama that is revealed as the audio-visual features. We investigated how these features put effect on viewers so that they feel a scene impressive. The result of our investigation would be a guideline for how audio-visual features are treated for effective movie digest.

１．はじめに

地上波放送に加えて各種衛星放送により数百チャンネルものコンテンツが提供される状況になっており，家庭の映像機器も従来の VHS カセット式のビデオレコーダがそのシェアを低下させ，数百時間分の映像を記録し，視聴可能なハードディスク内臓ビデオレコーダや録画可能な PC が普及している．このように，一般ユーザでもアクセスあるいは管理可能な映像コンテンツの量は増大しており，コンテンツ選択を効率的にするためのインタフェースの必要性が高まっているといえる．

映像コンテンツは時間の概念を持つため静止画像と比較すると一覧性が低いという問題がある．未視聴であるコンテンツの取捨選択タスクを効率的にするためには，映像内の特徴的なイベントなどを検出して構

造化することが必要となる．そして，コンテンツの構造を可視化してブラウジングする機能を提供したり，重要な部分を取り出してダイジェスト生成し，その閲覧を通して視聴するコンテンツを選択させるなどの方法が考えられる．ここで，映像における「重要な」場面とはどのようなもので，それをどのように抽出するかが問題となる．

映像における「重要な」場面，言い換えればいわゆる「見所」はコンテンツの種類に依存すると考えられる．例えばサッカーであればシュートが決まって得点が入るシーンや，ドリブルで相手プレーヤを抜いてゴールポストに接近する場面などであろうし，ニュース映像であればアナウンサーが事件の内容を説明する場面よりも現場のライブ映像のほうが視聴者の関心をより集めるであろう．これらの見所を判定するための手法として，スポーツ映像に対しては画像処理によりボールを検出して追跡し，ゴールポストを通過する状態を検出したり，ニュース映像では映像テンプレートによりキャスターがニュースを読み上げる映像とライブ映像とを区別し，映像を構造化する手法などが提案されている．

これらの手法で抽出している特徴は異なるが，ニュース映像やスポーツ映像に見られる映像構成や編集の定型性，あるいは重要である場面の定型性からその抽出が比較的容易なコンテンツであるといえ，その研究例も多い．それに対して映画やドラマといったコンテンツはニュース映像やスポーツ映像と比較して映像を構成する被写体の構成などにそのような定型性は無いといってよく，ニュースやスポーツ映像とは異なった「見所」判定手法が必要である．

映画やドラマといったコンテンツの選択支援をするためのダイジェスト映像生成を考えた場合，少なくとも 2 種類の目標が考えられる．1 つは当該コンテンツのあらすじが理解可能であるようなダイジェストを生成することで，もう 1 つは印象的な場面などコンテンツの感性的な情報，言いかえれば雰囲気を伝えるダイジェストを生成することである．映画は芸術的な作品でもあることから，後者の目標を満たすダイジェストが未知コンテンツの鑑賞を前提とした映像選択支援には有用であると考えられる．そこで本稿では後者の目標に沿ったダイジェスト生成を想定する．

印象的な場面であるかどうかの判断基準は個々の視聴者によってばらつきがあると考えられる．逆の立場である制作者がある場面において何らかの印象を強調しようする場合，特定の撮影，編集上の演出を施すことは広く行われている．従って，これを検出することで，少なくともコンテンツを提供する側が意図して創り上げた印象的な場面を検出することは可能だと考えられる．そこで，本稿では，映画の撮影，編集時に多用される演出技法を検出することにより，ある場面が印象を強調する意図を以って作られたものであるかどうかを判定することを考える．そのために，映画，ドラマで多用される撮影，編集上の代表的な演出技法であるカメラワーク，カット割，BGM がどのように，あるいはどの程度場面の印象付けに寄与するかを明らかにすることを試みた．また，感性情報の強調に用いられるカメラワークの緩急とそれによって視聴者が受ける印象強度についても調査し，これら技法の特性と感性情報の強弱への影響を調査し，映像選択を目的としたダイジェスト生成の指針を明らかにすることを試みた．

2．関連研究

動画は基本的に映像の編集点であるカット以外の部分では，フレーム間の類似性が高く，冗長であるため，映像が大きく変化した部分をキーフレームとして検出し，それに基づいて映像を要約するという考えがある[1]．しかしこの手法は動画像が基本的に持つ性質である映像の冗長性を低減する意味での要約にしかならず，映像の印象などの感性情報に基づくものとはなり得ない．内容を考慮した要約方法としては音声認識，あるいはメタデータとして得られるトランスクリプト情報に対して tf-idf などによる重み付けをして自然言語のドメインで重要性判別をし，重要語に対応する映像部分を要約映像として得る手法[2]もある．この手法はニュース等の映像に対してはコンテンツの性質上有効であるが，映画やドラマ等における場面の印象評価は困難であり，映像の特徴と統合し感性情報に関する重要性を判定することに関しては十分な検討がなされていない[3]．

映像や背景音（BGM，効果音）の心理的誘引性に着目した手法としては[4]がある．この手法では一定時間内のカット出現数，音声区間，BGM 区間の割合，効果音の有無といった統計量に基づき心理的重要度を決定しているが，これら特徴量に関するそれ以上の分析はなされておらず，印象の分類やそれに基づいたダイジェスト生成には課題が残り，カット頻度以外の，カメラワークなど映像上の演出と感性情報との関係に関しては言及していない．映像変化，カメラワーク，音量，音楽等の特徴に対するユーザの傾注モデルを定義し，それに基づいた要約手法も提案されているが[5]，各特徴量を映像要約に用いることの妥当性や視聴者への印象付けへの寄与の程度等に関しては経験則に基づいている．

3．撮影，編集における演出－映画の文法

映画等の制作者側が場面の印象を強調するために施す撮影，編集上の技法とそれにより強調される感性情報との関係についてはいくつかの文献で著されている[6-8]．ここではカメラワークやショット長遷移に関する技法とそれにより強調される感性情報との関係について述べる．

3.1 カメラワーク

感性情報を強調するカメラワークにはカメラの光学系の変化により被写体の撮影上の大きさを大きくさせるズームイン，その逆の操作であるズームアウト，カメラ自体を被写体に近づける操作であるキャラクタドリー，その逆の操作であるプルバックがある．それぞれの操作により強調される感性情報について以下にまとめる．

(1)速いズームイン

主となる被写体以外を除外する視覚的句読点となり，人物の反応を強調したり，緊迫感を与える効果がある．

(2)速いズームアウト

場面から押し出されて現実に引き戻されるような効果を与えたり，開放感を強調する．

(3)人物に対する遅いズームインあるいはキャラクタドリー

被写体である人物への共感を促し，感情・情緒面あるいは心理的啓示を強調する効果がある．

(4)人物に対する遅いズームアウトあるいはプルバック

被写体である人物の悲しみや孤独感を強

調，あるいは心理的に遠ざかるような効果をもたらす．

3.2 ショット長遷移による効果

連続した複数ショット（編集点であるカットからカットまでの間の映像区間）の時間長を特定の傾向を以って遷移させることにより感性情報を強調する技法であり，以下の 4 通りの技法があることが述べられている．

(1)短いショットの連続

短いショットを連続させることにより，激しい状況(アクション性)や慌しい状況を強調する．しばしば動きのある映像に対してこのような編集がなされる．

(2)徐々に短くなるショット

複数ショットを時間順に徐々に短くしていくことにより緊迫感を強調する．

(3)長いショットの連続

平穏な状況や落ち着いた雰囲気を強調する．

(4)徐々に長くなるショット

緊迫からの開放感を強調する．

4．演出技法と印象の関係

4.1 演出技法と印象付加の割合

撮影，編集時に制作者が場面の印象を強調するために用いるカメラワーク，ショット長遷移に加えて BGM を取り上げ，これらの特徴を含むショットに対して視聴者が印象を受ける度合いを調査した．

ここでは「ショーシャンクの空に」（フランク・ダラボン監督，ドラマ），「スパイダーマン」（サム・ライミ監督，アクション），「シャイニング」（スタンリー・キューブリック監督，ホラー）の 3 本の映画を大学生，大学院生計 10 名の被験者に鑑賞してもらい，各々の被験者が印象的であったと感じたショットを回答してもらった．

実験では，32 インチ液晶モニタの前に，水平方向の呈示画角が約 45 度となる距離（約 85cm）に座ってもらい，映画館のスクリーンに近い臨場感を得られるように視聴環境を設定した．1 回目は上記環境で映画を鑑賞してもらい，始めから終わりまで見終わった後に 2 回目の視聴を課し，印象的であると感じた場面をショット単位で挙げてもらった．なお，複数回視聴するうちに印象が変化する可能性を極力抑えるため，2 回目に同一映画を視聴しながら印象を受けた場面を挙げてもらう際には水平方向の呈示画角を約 10 度とし，臨場感を受けないよう配慮した．なお，ショット指示の際は映像とともにショット番号を常に表示するインタフェースを用意し，印象に残ったショットの番号を用紙に記入してもらった．なお，カメラワーク，ショット長遷移，BGM 等特徴量の具体的な検出手法は [9-11]で述べている．本稿で述べる実験のための正解データは，まず，これらの手法により特徴量を検出した後，目視等によりエラーを修正して得たデータとした．

各ショットを施された撮影，編集上の演出とその組み合わせから以下の 8 種類に分類し，各々の被験者が印象的であると感じたショットの傾向を分析した．

(1)カメラワーク
(2)ショット長遷移
(3)BGM
(4)カメラワーク＋ショット長遷移
(5)カメラワーク＋BGM
(6)ショット長遷移＋BGM
(7)カメラワーク＋ショット長遷移＋BGM
(8)カメラワーク，ショット長遷移，BGM な

し

ここでは，半数以上の被験者が印象的であると判断したショットを，多くの被験者に印象を与えるショットであると判断し，その割合を調べた．ここで，半数以上の被験者が印象的であると判断した上記(1)から(8)のうちのいずれかの演出(の組み合わせ)がなされたショット数 Ic を同じ性質を持つショットの総数 Rs で割った値(Ic=Is/Rs)を演出の印象強度として，各々の種類のショットに対してその値を算出した．また，カメラワーク，ショット長遷移，BGM による演出の有無を問わず，半数以上の被験者が印象的であると判断した場面のショット数を Ai として，上記(1)から(8)の演出が施された区間のうち印象的であると判断した区間が Ai に占める割合を印象貢献度 Or=Ic/Ai としてその割合を求めた．その結果を表１から表４にまとめる．なお，表において CW：カメラワーク，CT：ショット長遷移，BGM：バックグラウンドミュージックである．また，"none"はこれら３種類の演出が施されていなかったショットを表す．

ジャンル毎の映画のサンプル数が少ないため今後サンプル数を増やして検証する必要があるが，表 1 から表 3 の Ic の値より，カメラワーク， ショット長遷移，BGM いずれかが単独で適用された場合，それらの約半数のショットが印象的なものであることがわかった．カメラワークの施されたショット数はショット長遷移や BGM と比較して少ないが，印象強度としては他の 2 つと同程度であるので，映像の印象付けに関してこれらと同様な効果があることがわかる．また，この Ic の値は 1 つのショットに対して同時に施される演出技法の数が増えるにしたがって高くなる傾向にあり，3 つの演出

表１　「ショーシャンクの空に」の結果

	CW	CT	BGM	CW+CT
Ic	0.52 (47/91)	0.76 (127/167)	0.48 (216/452)	1.00 (2/2)
Or	0.12 (47/394)	0.32 (127/394)	0.55 (216/394)	0.01 (2/394)

	CW +BGM	CT +BGM	CW+CT +BGM	none
Ic	0.54 (26/48)	0.89 (107/120)	1.00 (1/1)	0.23 (120/517)
Or	0.07 (26/394)	0.27 (107/394)	0.00 (1/394)	0.30 (120/394)

表２　「スパイダーマン」の結果

	CW	CT	BGM	CW+CT
Ic	0.49 (40/81)	0.47 (273/581)	0.39 (310/785)	1.00 (6/6)
Or	0.08 (40/511)	0.53 (273/511)	0.61 (310/511)	0.01 (6/511)

	CW +BGM	CT +BGM	CW+CT +BGM	none
Ic	0.74 (35/47)	0.53 (163/309)	1.00 (6/6)	0.12 (79/649)
Or	0.07 (35/511)	0.32 (163/511)	0.01 (6/511)	0.15 (79/511)

表３　「シャイニング」の結果

	CW	CT	BGM	CW+CT
Ic	0.51 (43/85)	0.34 (70/205)	0.46 (133/289)	0.54 (15/28)
Or	0.20 (43/219)	0.32 (70/219)	0.61 (133/219)	0.07 (15/219)

	CW +BGM	CT +BGM	CW+CT +BGM	none
Ic	0.48 (25/52)	0.40 (33/82)	0.50 (7/14)	0.32 (32/101)
Or	0.11 (25/219)	0.15 (33/219)	0.03 (7/219)	0.15 (32/219)

表４　３本の映画の平均値

	CW	CT	BGM	CW+CT
Ic	0.51	0.52	0.44	0.85
Or	0.13	0.39	0.59	0.28

	CW +BGM	CT +BGM	CW+CT +BGM	none
Ic	0.59	0.61	0.83	0.22
Or	0.08	0.25	0.15	0.20

技法が併用されているショットは高い率で視聴者の印象に残る場面となっていることがわかった．つまり，これら異種特徴量の組み合わせと感性情報の強弱は相関傾向にがあることがわかり，印象的な映像をスキミングして映像選択支援のダイジェスト生成をするためにはある場面で施されている撮影，編集上の技法の数に応じて重み付けをすれば良いことがわかった．

4.2 カメラワーク速度と印象強度の関係

次に，カメラワーク速度の違いが視聴者に与える印象内容や強度にどう影響するかを検証する実験を行った．実際の映画の映像から異なるズーム速度の映像を用意して被験者に視聴させても，被写体等が異なれば，それが原因となって異なる印象を受ける可能性が排除できないため，ズーム速度を計算機制御可能なカメラにより変化させて同一被写体，同一背景下で撮影し，それを被験者に視聴させ印象を評価してもらった．

まず，2名の人物各々に対して8段階の異なる速度でズームイン／ズームアウトさせながら撮影した映像を用意した．ズーム速度以外の要素は極力同一となるよう配慮した．次に 8 通りの速度のズームイン映像映像×2 名＝16 通りの映像をランダムに提示し，印象の強弱に関するアンケートを行った．ズームアウト映像に関しても同様な手順で実験を行った．被験者は大学，大学院学生計 4 名であり，先の実験と同様水平方向の呈示画角が 45 度となるように被験者を配置した．ズーム速度は，映像フレームの対角線を定線として得られる時空間投影画像[11]に現れる被写体のエッジの軌跡が時間軸に対してなす角(deg./sec.)により表すこととする．この値の絶対値が大きいほどズーム速度が速いことを意味する．ズームインの速度は 2～30deg./sec.（被写体の拡大率にして5秒間に約 1.4 倍から 8.4 倍）の範囲で8 段階の速度を設定し，ズームアウトの速度は-11~-36deg./sec.(5 秒間に約 1/4.2 から 1/16.4 倍に縮小)の範囲で 8 段階の速度を設定した．

3.1 で述べたように，ズームイン映像であれば緊迫感や心理･情緒面の強調，ズームアウト映像では開放，孤独感に関する印象についてそれらが強調されているか否かを 5 件法により，回答してもらった．この際，印象が強調されていると強く感じるならば+2，印象が強調されていないと強く感じるならば-2，どちらでもないならば0の評点をつけることとし，全被験者の平均評点を求めた．

ズームイン，ズームアウト映像に関する印象評価の結果を図 1，図 2 に示す．図 1 より，ズームイン映像に関しては，ズーム速度が比較的速い約 8~30deg./sec.(5 秒間に被写体が約 2.1～8.4 倍拡大)であるような映像ではズーム速度と緊迫感に関する印象の強弱の間に正の相関があり，ズーム速度が約 2~5deg./sec.(5 秒間に約 1.4～1.7 倍拡大)であるような緩やかなズームインではカメラ速度と心理･情緒面についての印象の強さに負の相関があることがわかった．

一方，ズームアウト映像に関しては，ズーム速度が比較的速い約-20~-36deg./sec.（5 秒間に約 1/4.2～1/16.4 に縮小）でズーム速度と開放感に関する印象との間に正の相関があり，ズーム速度が比較的遅い-11～-18deg./sec.（5 秒間に約 1/2.3～1/3.3 に縮小）ではズーム速度と孤独感の印象との間に負の相関があることがわかった．

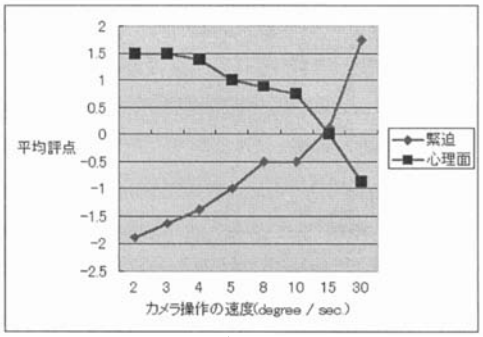

図1　ズームインの速度と印象の関係

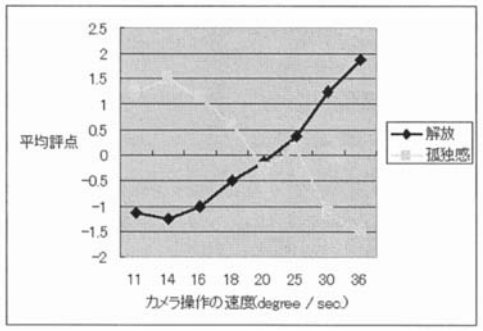

図2　ズームアウトの速度と印象の関係

5．　まとめ

本稿では，映画やドラマなどの撮影・編集上の技法により感性情報が強調される性質を持つコンテンツと対象とし，映像から受ける印象に基づき未視聴コンテンツの取捨選択を助けるためのダイジェスト生成に着目した．そして，ダイジェスト生成のための特徴量の評価に関する実験結果について述べた．現在までの映像要約に関する研究ではカット頻度やカメラワーク，あるいはBGMの有無等を映像の重要度評価に用いることが経験的に提案されているものの，これらの特徴量が視聴者の印象付けにどの程度寄与しているかに関しては明らかにされていなかった．

本稿で示した実験の結果より，解析した映画の数がまだ不十分ではあるものの，カメラワークによる演出はショット長遷移やBGMによる演出と同程度の印象強調効果があり，より多くの技法が同一映像区間に施されているほど印象強調効果も高まることがわかった．この結果から，映像の印象に基づくコンテンツ選択を目的としたダイジェスト生成における映像の重み付けに対する指針が得られたと考えている．

また，ズーム操作に伴って強調される感性情報の強調効果が切り替わる速度があることが明らかになり，異なる感性情報を強調する同一のカメラ操作がどのような速度を境界としてその効果を分岐させることになるかが明らかになった．この知見を，特定の感性表現にバイアスをかけたダイジェスト生成等へ応用することなどが考えられる．

今後の課題としては，本稿では言及しなかった，キャラクタドリーやプルバックに関しても同様な分析を進め，カメラ操作に関連する印象強調効果をさらに明らかにすることが考えられる．

謝辞

本研究の一部は科学技術振興機構「地域イノベーション創出総合支援事業シーズ発掘試験」による助成を受けた．ここに記して謝意を表す．

参考文献

[1] G. Ciocca, R. Schettini, “Dynamic key-frame extraction for video summarization”, Proc. Internet imaging VI, Vol. SPIE 5670, pp. 137-142, 2005.

[2] 伊藤一成，酒井康旭，斎藤博昭，”音声と映像の一貫性を考慮した要約動画の生成”，電子情報通信学会・データベース学会合同

ワークショップ DEWS, 2004.

[3] Michael A. Smith, Takeo Kanade, “Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques”, IEEE Computer Vision and Pattern Recognition, pp. 775-781, 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, 2001.

[5] Yu-Fei Ma, Lie Lu, Hong-Jiang Zhang, Mingjing Li, “A User Attention Model for Video Summarization”, Proc. of ACM Multimedia, pp. 533-542, Dec. 2002.

[6] 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.

[7] ジェレミー・ヴィンヤード著, 吉田俊太郎訳, “傑作から学ぶ映画技法完全レファレンス”, フィルムアート社, 2002.

[8] ルイス・ジアネッティ著, 堤和子, 増田珠子, 堤龍一郎訳, “映画技法のリテラシー I 映像の法則”, フィルムアート社, 2003.

[9] 川崎智広, 吉高淳夫, 平川正人, 市川忠男, “映画における音楽, 効果音の抽出及び印象評価手法の提案”, 信学技報, MVE97-96, pp. 23-29, 1998.

[10] A Yoshitaka, T. Ishii, M. Hirakawa, and T. Ichikawa, ”Content-Based Retrieval of Video Data by the Grammar of the Film,” Proc. of International Symposium on Visual Languages, pp. 310-317, 1997.

[11] 吉高淳夫, 松井亮治, 平嶋宗, ”カメラワークを利用した感性情報の抽出”, 情報処理学会論文誌, Vol. 47, No. 6, pp. 1696-1707, 2006.