Vol.2015-MUS-107 No.18
2015/5/23

音声と映像の変化に注目したフレーム間引きによる動画要約手法

平井 辰典[1,a)]　森島 繁生[2,3]

概要：本稿では，動画を短時間で視聴することを目的としたフレーム間引きによる動画要約手法を提案する．本要約手法では，動画における音声と映像の変化に注目して，変化の少ない冗長なフレームを間引いていくことで動画の内容を保持したまま動画の長さを短くする．フレームを間引くごとに，間引いた箇所の音声と映像の変化を再計算することで，動画の連続性を補償したままの要約が可能となる．本手法によって，より短時間で動画を視聴することができるだけでなく，変化の少ないフレームを挿入することで音声や映像がスロー再生されることなく動画の長さを伸ばすことも可能となる．主観評価実験によって本手法の有効性について検討した．本稿ではさらに，本手法を応用することによる動画中の移動物体の削除や音楽と映像の同期についても検討する．

1. はじめに

テレビをはじめ，DVD やインターネット，個人の携帯端末など，動画コンテンツはいたるところに存在している．多くの動画はすでに完成されたコンテンツであり，我々はあるがままに鑑賞している．従来，動画コンテンツはあるがまま鑑賞されるものであり，我々は動画の再生時間に合わせて鑑賞に充てる時間を決めていた．一方で日常生活で動画コンテンツの鑑賞などの趣味に充てられる時間は有限であり，その鑑賞可能な時間が動画の再生時間と一致することは稀である．本研究では，再生時間に合わせて鑑賞をするのではなく，鑑賞可能な時間に合わせて動画の内容を変えずに再生時間を可変にする手法を検討する．

動画共有サービスの一般社会への浸透と共に動画コンテンツの数は膨大化しており，インターネット上には個人が生涯鑑賞しきれないほどの動画コンテンツがアップロードされている．テレビの地上波放送は 1953 年の放送開始以来，複数の局で番組が放送され続けている．このように単調増加を続ける動画コンテンツの中から，興味のあるコンテンツを効率的に鑑賞するための手法は注目を集めている．例えば，動画の検索や推薦などの鑑賞すべき動画を決定するためのアプローチや動画を短時間で鑑賞するための動画要約手法など，様々な側面から研究が行われている．特に，従来はメモリ等の制約から自由に扱うことのできなかった動画コンテンツは，近年の計算機の発達に伴いより自由度を増しており，研究対象としても注目を集めている．

本稿では，動画を短時間かつ効率的に視聴することを目的とした動画要約手法に関する新たなアプローチを提案する．動画要約の目的は，動画の内容を理解するために必要な情報をなるべく削らず，単位時間当たりに得られる情報量を最大化することである．ただし，単位時間当たりの情報量をただ増やすだけではなく動画の内容を理解可能な状態に保つ必要がある．本研究では，動画における音声と映像の変化に注目して，変化の小さい箇所をフレーム単位で間引いていくことで，動画の内容を保ったまま要約する手法を提案する．これにより，動画の内容やフレーム間の連続性を保ったままその冗長性のみを削減する．

また，本研究では動画における音声と映像の変化が小さい箇所にフレームを挿入することで，動画の伸長も実現する．動画要約手法は，動画の鑑賞支援という観点で重要な技術であるが，我々は動画の要約（収縮）と伸長によってその再生時間を可変にし，動画編集の支援も目指す．動画編集では，放送時間などの制約から決められた再生時間に対して，動画がフィットするために最適な編集点 [*1] を試行錯誤しながら探すことで編集を行うことが多い．仮に撮影した素材が決められた再生時間に完全にフィットする場合，編集の手間は少なくなると考えられる．本研究では，動画の再生時間を可変にすることで，動画の鑑賞や編集といった様々な用途で動画を扱う上での効率化を目指す．

[1] 早稲田大学
Waseda University, Shinjuku, Tokyo 169–8555, Japan
[2] 早稲田大学理工学術院総合研究所
[3] JST CREST
[a)] tatsunori_hirai@asagi.waseda.jp

[*1] 動画をカットする際の境界となる点

2. 関連研究

動画要約は動画研究における重要な分野の一つに位置付けられており，様々なドメインに関して研究されてきた [1]．特にスポーツ動画に関する動画要約は盛んに行われており，河村らはラケットスポーツ動画において重要なラリーシーンを抽出することで動画を要約する手法 [2] を，Tjondronegoro らはホイッスル音や観客の盛り上がり，テキスト情報を基にスポーツ動画のハイライトシーンを検出することで動画を要約する手法 [3] を提案している．

DeMenthon らは，動画を多次元特徴量空間中の曲線で表現し，曲線の単純化によってキーフレームを検出し，動画を要約する手法を提案した [4]．この手法ではフレーム単位で動画の間引きを行い，ユーザが指定した任意の長さに動画を要約することを可能としているが，音声については考慮していない．Smiths らはシーンチェンジやカメラの動き，物体認識の結果，音声中のキーワードを統合した動画の内容に関する指標を基にして，動画セグメントを間引くことにより内容を保持したままの動画要約手法を提案した [5]．この手法では，動画の内容を理解する上で重要でない箇所のみを削ることで動画の内容を保持するが，シーンが削られてしまうため，動画の元々の連続性は保たれない．つまり，要約結果は編集後の動画と同等のものとなる．

動画要約手法は多くの場合，動画の重要な箇所をピックアップするアプローチか重要でない箇所を削るというアプローチに分けられる．その一方で，動画の内容を基にシステムがその一部を削る手法では，視聴者は何が削られてしまったのかがわからないため，動画を削らずに短時間での視聴を可能とするための高速再生手法が提案されている．栗原らは，動画の音響解析を基に，動画を発話区間と非発話区間に分け，発話区間では音声を聞き取れる速さで動画を高速再生し，非発話区間では動画中の動きが把握可能な速さで高速再生することで，動画を短時間で視聴することを可能とした [6]．この手法は音声に特化した手法であり，映像の内容を考慮しないため，スポーツ動画など必ずしも音声が重要ではないような動画には適していない．Peker らは映像中の動きに応じて [7]，Cheng らはユーザの興味に応じて [8]，伊藤らは映像の変化度合いに応じて [9] 動画の再生速度を可変にする手法を提案したが，これらの手法では音声を扱っていないため，対話動画など，音声が重要な役割を担うような動画には適していない．再生速度を変換するアプローチは動画以外のメディアに対しても適用されている．都木らは，ラジオ放送などの音声放送において音声全体の時間は伸ばさずに，発話区間を伸長し非発話区間を収縮することで音声を聞き取りやすくするための話速変換技術を提案している [10]．

本稿では，動画要約において動画の内容を保持するために，フレーム単位で動画を間引くことによる要約手法を提案する．フレームを間引く上で注目するのは，音声及び映像の変化である．単に変化の小さいフレームを間引くのではなく，間引いた後で新たに生じるフレーム間の変化も小さくなるようなフレームを間引く．本手法では，前後フレーム間の変化に注目しているため，動画の時間的な連続性が保持される．それにより，動画要約手法のように要約結果に編集点が生じにくく，ありのままに近い形で動画の再生時間を変化させられる．これは動画の再生速度を可変にするアプローチに近く，フレームを間引くことは局所的に高速再生をすることと同義であるが，我々は本手法を鑑賞という観点だけではなく編集にも適用可能であることを明確化させるため，高速再生ではなくフレーム間引きによる動画要約手法と呼ぶ．また，本手法では，音声と映像の双方を考慮したフレーム間引きを行うことによって，前述した動画の高速再生手法では考慮できていなかったマルチモーダルな動画の伸縮を実現する．

動画要約や高速再生手法などでは，動画の内容理解や効率的な再生など，鑑賞を支援することを目的としているが，我々は鑑賞のみではなく，動画編集の支援も目指す．動画編集において，使用する映像素材や音声素材の長さがありのままですべて一致することはほとんどない．そのため，編集者は映像と音声が決められた再生時間に収まるように試行錯誤によって動画を編集する．佐藤らは動画の長さに合った BGM を付加することを目的に，楽曲を小節毎に分割して，映像の決められた箇所と楽曲の決められた箇所が同期するように楽曲そのものを切り貼りによって再構成する手法を提案した [11]．これによって，付加したい楽曲の長さが映像と一致しない場合でも動画に BGM を付加することを可能とした．Berthouzoz らはインタビュー動画の編集を目的として，ユーザが使用する箇所を選択し，選択された箇所同士をシームレスに遷移させることで一連の連続的なインタビュー動画の生成を実現するシステムを提案している [12]．この手法では，視聴者に編集点が気づかれないように動画の長さを変更することを可能としているが，インタビュー動画のように被写体があまり動かず，カメラと被写体の構図に変化が少ないような動画にのみ適用可能である．本研究では，動画の再生時間の伸縮というアプローチによって動画編集の支援を目指す．再生時間を可変に実現するために，動画のフレームを間引くだけではなく変化の小さいフレームを挿入することによって動画を伸長する手法についても提案する．

3. フレーム間引きによる動画要約

動画は，フレームを時々刻々と切り替えていくことで単独の画像では静止しているものを動いているように見せるメディアである．一般的に普及している動画は 1 秒間に約 30 枚のフレームからなり，フレームとフレームの間隔は約 0.033 秒である．1 秒当たりのフレーム数をフレームレー

トと呼び，テレビなどの一般的な動画は 29.97FPS（frames per second）である．実際には動きを絶え間なく連続に撮影しているわけではないが，フレームとフレームの間に存在するはずの視覚的情報を脳が補間するため，動画では物体が動いているように見える．多くの映画やアニメは 24FPS であったり，Web にある動画の一部は，そのサイズを抑えるために 15FPS で保存されていたりもする．フレームレートが高いほど動きが滑らかに，フレームレートが低いほど動きが角ばって見えるが，ある程度のフレームレートがあれば人間の脳は動きを補間して捉えてくれる．

30FPS の動画のフレームを一枚おきに間引いて再生すると 2 倍速再生となる．このとき，動画の情報量 [*2] は 15FPS で撮影した際と同等となる．ここで，動画における映像の変化に注目する．例えば，静止した物体を 30FPS で撮影した動画のフレームを半分に間引いたとしても，動画から得られる視覚的な情報の量は変わらない．極端に言うと，完全に静止している物体であれば 1 フレームのみ残して他のフレームすべてを削ったとしても得られる情報は変わらないのである．このように，動画における映像の変化とそこから得られる情報量に注目すると，変化のないフレームを間引いても動画から得られる情報量は変わらない．

本研究の根幹となるアイデアは，動画における変化に注目して変化の小さいフレームを優先して 1 フレームずつ間引いていくことで，情報量は削減せずに（動画の内容を保持したまま）動画の再生時間を短くするというものである．音声についても同等のことがいえ，まったく同じ音が鳴り続けている場合，その一部が間引かれていても内容は変わらない．これを踏まえて本稿では，音声と映像の変化に注目したフレーム間引きによる動画の要約手法を提案する．

3.1 映像の変化に基づくフレーム間引き

人間の視覚は急な変化に敏感なため，動画においてフレーム間の時間的な連続性は重要な要素である．同じ 1 フレームを間引く場合にも，フレームの内容が大きく変化する箇所の 1 フレームと変化が小さい箇所の 1 フレームではその違和感の差は顕著である．例えば，高速に移動する物体を固定したカメラで撮影する場合，物体の移動距離は大きく，限られたフレームでしか物体を撮影できない．このような場合に，物体を捉えた 1 フレームを間引いてしまうと，物体の動きの連続性に関する情報を失ってしまう．一方で，物体の移動速度が遅い場合には，フレームを間引いても，その移動距離が小さいため，移動に関する連続性の情報はあまり失われない．

以上のことから，我々は物体の動きを総括的に扱えるフレーム間の変化に注目する．具体的には，フレーム間の画素値の差分である Sum of Squared Differences（SSD）に注目する．フレーム t における SSD を表す $s(t)$ は動画の画面サイズ $width$, $height$ と画素値 f より以下の式で表せる．

$$s(t) = \sum_{y=1}^{height} \sum_{x=1}^{width} (f_t(x,y) - f_{t-1}(x,y))^2 \qquad (1)$$

[*2] 本稿では，動画から得られる視覚的情報量のことを情報量と呼ぶ．

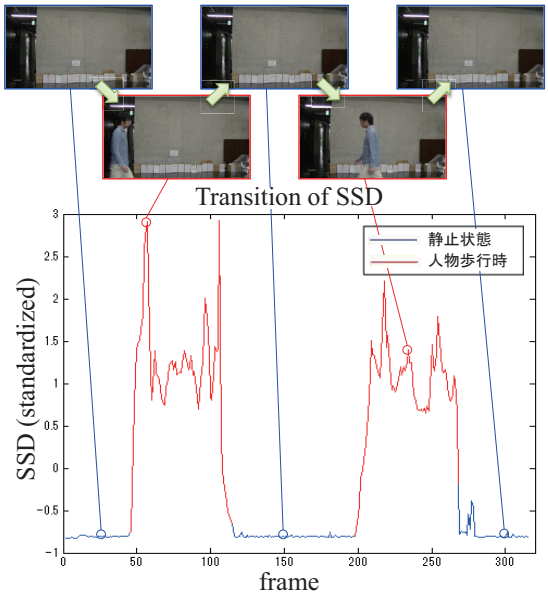

図 1 動画における SSD の遷移
Fig. 1 Transition of SSD in a Video Clip.

固定したカメラの前を人物が 2 度横切る動画における SSD の遷移を 図 1 に示す．ここで，動画における重要なイベント（人物の歩行）において SSD の値が大きくなっていることがわかる．

このような動画で SSD が小さいフレームから順にフレームを間引くと，物体が動いていないフレームが優先的に間引かれていく．このように，SSD の値でソートして，値の小さいフレームから間引いていくことで，動画の内容を保持したまま動画の再生時間を短くすることができる．

この手法で，ある程度はうまくフレームを間引くことができるが，実際には，フレームを間引くことによって新たに変化が生じてしまう．本手法では，フレームを間引く際に新たに生じる変化も考慮する．具体的には，該当フレームの SSD の値と，そのフレームを間引いた際に生じる新たな前後フレームによる SSD の値の和をフレーム間引きのためのコスト C_{video} として，コストが最小なフレームから順にフレームを間引いていく．これによって，映像全体の連続性がなるべく保持されるようなフレームの間引きが可能となる．このようなフレームの間引きを行うために，1 フレーム間引くたびに該当フレームの前後のフレームに関するコストを再計算する．ここで，再計算した値と標準化された特徴量の値との間には差が生じるため，標準化の際に使用する平均と標準偏差の値を保持しておき，それらを用いてスケールを調整する．また，1 フレーム目と最終フレームに関しては間引かないものとする．

Avidan らによる Seam Carving という画像の内容を保持したままのリサイズ（リターゲッティング）手法では，隣接

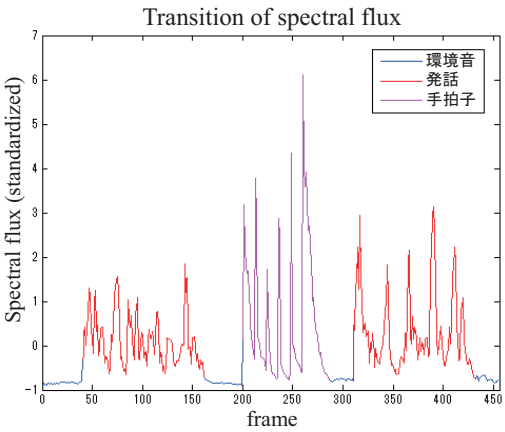

図 2　音声中のスペクトルフラックスの遷移
Fig. 2　Transition of Spectral Flux.

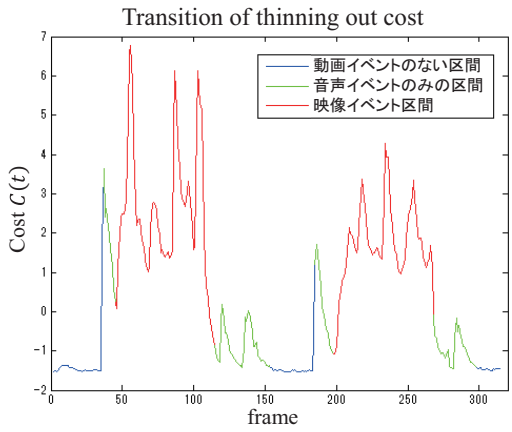

図 3　動画フレームを間引くためのコストの遷移
Fig. 3　Transition of Thinning Out Cost.

ピクセル間の変化が小さいパスを探索し，画像のピクセルを 1 列毎に間引く [13]．これを発展させた形で Rubinstein らにより提案された Improved Seam Carving では，隣接ピクセル間の変化に加えて新たに生じる変化も小さくなるようなパスを探索し，ピクセルを 1 列毎に間引く [14]．本手法は Improved Seam Carving を動画のフレーム間の関係に応用し，リターゲッティングではなく動画要約を実現する手法であるといえる．

3.2　音声の変化に基づくフレーム間引き

音声は 1 秒当たりの標本数 44,100 や 48,000 などと，動画における 1 秒当たりのフレーム数と大きな隔たりがある．そこで，音声を処理するにあたって動画と同じ時間幅で分析できるように，音声を音声フレームに分ける．動画のフレームレートを r としたとき，音声のフレーム長は $2/r$ 秒とし，$1/r$ 秒毎にシフトさせながら分析を行う．これにより音声と映像それぞれのフレームが時間的に同期する．

音声の変化を表す特徴量としてスペクトルフラックスを用いる．スペクトルフラックスは音声スペクトルの局所的な時間変化を表すものであり，音のオンセットやオフセットなど，音に変化が生じる箇所で特に高い値を示す特徴量である．スペクトルフラックスの抽出には Lartillot らによる音楽信号解析ツールである mirtoolbox1.5 を用いた [15]．

図 2 に発話，拍手時のスペクトルフラックスの値の遷移の様子を示す．音声の収録環境は室内環境で，上述以外の大きな音声イベントは発生していない． 図 2 から，このようなシンプルな状況であれば，スペクトルフラックスに注目することで音声イベントが起こっている箇所が検出できることがわかる．

スペクトルフラックスの値が小さい音声フレームから順に間引くことで，音声イベントの内容を失うことなく音声を短くすることができる．ここでも 3.1 節での映像フレームの間引きと同様に，フレームを間引くことによって生じる新たな変化も考慮したコスト C_{audio} を設定し，逐次計算しながらコストの小さいフレームを間引いていく．

3.3　音声と映像の変化に基づくフレーム間引き

映像の変化に注目したフレームの間引きでは，映像的に変化が少なく，間引いても連続性が保たれるようなフレームから順に間引いてきた．また，音声の変化に注目したフレームの間引きでは，音声の連続性を保持しつつ音声イベントを残すようなフレームの間引きを行ってきた．動画を要約する上で，我々は両方の要素を考慮する．そこで本手法では，映像と音声のフレーム t を間引く際のコストをそれぞれ $C_{video}(t)$, $C_{audio}(t)$ として以下の式によって間引くべきフレームを決定する．

$$C(t) = \alpha C_{video}(t) + (1-\alpha) C_{audio}(t) \qquad (2)$$

ここで，α は音声と映像のどちらを重視するかを決める重みであり，本研究では音声と映像の重みを同等のものとするため，実験的に $\alpha = 0.5$ としている．また，α による重みの効果を均等にするために SSD 及びスペクトルフラックスの値は動画全体で事前に標準化している．式（2）によって表されるコストが最小となるフレームを逐次間引いていくことによって動画の要約を実現する．**図 3** に 3.1 節の図 1 で用いた歩行動画に対して算出したフレーム間引きのためのコストの遷移を示す．この動画では人物が横切る前後で足音のみが聞こえる箇所があるが，そのような箇所でもコストは高くなっており，映像中の動きだけでなく音声の変化も反映したコストとなっていることがわかる．

4.　フレームの挿入による動画の伸長

本研究では，動画の再生時間を可変にするために，動画の要約だけではなく伸長も実現する．3 章で述べた手法によって動画の要約は可能であるが，同じ特徴量を用いたコスト計算に基づいてフレームを挿入していくことで，動画の内容を保持したままで動画の再生時間を長くすることもできる．動画の伸長の実現のために SSD とスペクトルフラックスの重み付き和によるコストが小さいフレームを同一箇所に挿入する．ここで，要約手法とは違い，フレームを挿入することによって生じる新たな変化は考慮しない．

なぜならば，すでに存在するフレームと同じフレームを挿入するため，新たに変化は生じないためである．つまり，フレームを挿入する際にはコストの逐次計算はせず，コストをソートした順番にフレームを挿入する．

また，コストの小さいフレームから順番にフレームを挿入していくと，ちょうど二倍の長さにしたところで，すべてのフレームが二枚ずつ存在することとなってしまい，0.5倍の速度で再生していることと同義となってしまう．スロー再生をすることで動画の内容は保たれるが，物体の動きも一様に 0.5 倍となってしまい，物体の動きに関する情報が変化してしまう．そこで本研究では挿入に使うフレームの割合に上限を設定し，一定枚数のフレームを挿入したら，再度コストの小さいフレームから順に挿入することで物体の動きまでスローにしてしまうことを避ける．この上限の値は動画毎に違った数値であり，動画フレーム全体のうち内容には影響しないフレームの割合となるべきであるが，現状の実装ではユーザが定める値となっている．

5. 主観評価実験

本手法によって伸縮させた動画を視聴した際に，実際に伸縮された動画であると感じられるかを主観評価実験により簡易的に検証した．

7 名の男性被験者に対し，2 種類の動画に関して 50%，75%，100%，125%，150% のフレーム数となるような伸縮率で本手法を適用した動画，計 10 本を視聴してもらい，「1 : 動画が収縮されていると感じた」,「3 : 自然であった」,「5 : 動画が伸びていると感じた」という 5 段階で評価してもらった．ここで，音声と映像の重み α は 0.5，伸長する際に挿入するフレームは全体の 20% のフレームとした．さらに，比較のために同じ伸縮率で動画を一様に伸縮した変速再生動画 10 本についても視聴，評価してもらった．被験者には各動画について，提案手法による伸縮動画 5 本を視聴してもらい，その後変速再生動画 5 本を視聴してもらった．動画の提示順はランダムであり，被験者にはどの動画が元の伸縮なしの動画であるかがわからないようにした．ここで，被験者には前後に視聴した動画を踏まえた相対的な評価ではなく，各動画に対して独立な評価を心がけるように指示した．2 種類の動画はそれぞれ「No.1 身振りを交えたスピーチ動画（通常再生時間 22 秒）」,「No.2 ナレーションの後に花火が打ち上がる動画（通常再生時間 25 秒）」である．伸縮率 100% の動画は元の動画と同一のもので，提案手法及び変速再生で同一であるが，評価のばらつきを検証するために，重複して視聴してもらった．

主観評価実験の結果を**表 1** に示す．表 1 の評価値は 7 人の被験者の平均値であり，最下行の平均値は No.1 及び No.2 の評価値の平均値である．評価値が 3 に近いほど，動画の伸縮が感じられなかったことを示す．この結果から，伸縮率 75%，125%，150% において，提案手法の評価値の平均の方が 3 に近く，伸縮を感じづらい自然な伸縮を実現できていることがわかる．伸縮率 50% の動画については，提案手法でも収縮を感じられる結果となった．これは，提案手法の収縮において，映像及び音声の変化が少ないフレームがすべて間引かれており，変化の大きい重要なフレームまでも間引かれてしまったためであると考えられる．被験者から取ったアンケートの結果には，特に発話の伸縮や映像の不連続さが判断の要因になったとの意見があった．本評価実験は簡易的なものであったが，さらに多くの動画や伸縮率を扱った場合に評価結果がどのように変化していくかを基に，提案手法で違和感なく視聴が可能な伸縮率の範囲とその有効性に関して今後さらに踏み込んだ評価を行っていきたい．また，音声と映像に関する適切な重み α についても今後検証を行っていきたい．

表 1　主観評価実験の結果
Table 1　The Result of Subjective Evaluation Experiment.

動画	手法	伸縮率 50%	75%	100%	125%	150%
No.1	変速再生	1.00	**1.86**	3.00	4.14	4.86
	提案手法	1.00	1.71	3.00	**4.00**	4.86
No.2	変速再生	1.14	1.86	3.00	4.00	4.43
	提案手法	1.14	**2.29**	2.86	**3.71**	**4.00**
平均値	変速再生	1.07	1.86	3.00	4.07	4.64
	提案手法	1.07	**2.00**	2.93	**3.86**	**4.43**

6. 本手法の応用

本手法の応用による動画編集の支援について検討する．3 章では，変化の小さいフレームを間引いていくことによって動画そのものの要約を実現したが，逆に変化の大きいフレームを間引くことで移動している物体を削除することを検討する．4 章と同様のコストを用いて変化の大きいフレームを間引くことで移動物体は削除可能であるが，削除したフレームの前後における連続性も確保しなくてはならない．よって，カメラが静止している場合には，移動物体が映っている箇所がすべて削除された場合にも前後の区間の連続性は比較的保たれるが，カメラが動いている場合や複数の物体が複雑に動いている場合などには，本手法単体では移動物体の削除は困難である．そこで，該当物体を含む領域をアノテーションし，その前後のフレームを含めて動画の部分領域を定義し，部分領域内で変化の大きい箇所を間引き，さらに元の長さに戻るように，間引いた後の部分動画における変化の小さい箇所を挿入することで，移動物体を削除することが可能となる．しかし，この手法でもカメラの移動が大きい場合や移動物体そのものが大きい場合，広範囲にわたって移動する場合には，処理をした部分領域と元の動画の境界に違和感が生じてしまう．本手法では，カメラの動きがない場合や，移動物体が少ない場合であれば，ある程度の品質で物体を削除可能である．

本手法のさらなる応用として，動画における音と映像の

同期について検討する．3，4 章では音声と映像を同時に考慮してフレームの間引き及び挿入を行ってきたが，これらを別々に伸縮させることによる，動画における音と映像の同期の実現可能性を検討した．映像に対して音楽を後から付加するような動画編集において，音と映像の同期は重要な要素である．本手法では，任意の音楽と映像のセットを用意し，音楽の変化と映像の変化が同期するように映像を伸縮することで音楽と映像の同期を実現する．音楽は構造や発音時刻に関する制約の大きいメディアであるため，音楽の伸縮は行わなず，映像の方を伸縮する．与えられた楽曲のスペクトルフラックスに対し，映像の変化が同期するように，スペクトルフラックスと SSD を DP マッチングによって同期させる．これにより，大局的に音と映像の同期が実現できる．ただし，DP マッチングでは波形のピークの数が異なる場合に，本来同期させたいはずのピーク同士の適切なマッチが図れない．そのため，本手法は楽曲のある一部分と，映像の一部分を同期させるといった音と映像の同期を実現することに適していると考えられる．

映像を伸縮させることによって音声と同期させる手法については，飯塚らによって提案されており [16]，その有効性も示されているが，本手法では連続性に注目した音声と映像の同期を検討しており，具体的には注目する特徴量が違う．本手法による音と映像の同期の有効性については別途検証の必要があるが，本手法の応用の一つとして音と映像の同期が実現する余地があり，編集支援に関して広範囲で有効な手法となることが期待できる．

7. まとめ

本稿では，動画中の音声と映像の変化に注目してフレームを間引くことによる動画要約手法及び，動画の伸長や移動物体の削除，音と映像の同期を実現する手法を提案した．

本手法では，映像における「間」のように静止していること自体が意味を持っている場合にもフレームを間引いて動画を短くしてしまうため，時には動画の内容を変えてしまうかもしれない．そのように，変化しないことが意味を持つような場合には本手法は適さない．本手法は動画の内容を必ずしも意味的にも担保するわけではなく，視聴覚的に得られる情報を保持するものである．そのため，音楽のようにすべてのフレームが音響的に意味を持ったメディアに対しての適用はできない．また，連続性を保ったままでの要約可能な長さには動画毎の限界値があり，極端にフレームを間引くと，視聴覚的に重要なフレームも間引くことになってしまう．現状では，動画毎の要約可能な長さについては結果を視聴しながら確かめなければならないが，今後，伸縮可能な再生時間の範囲を自動で推定する手法についても検討したい．それとともに，本手法の有効性に関してより詳細な評価も行っていくつもりである．

動画鑑賞のための時間は有限である．渡邊らはその限られた時間に視聴すべき動画コンテンツを空いた時間に詰め込むことで効率的にコンテンツを消費するための手法を提案した [17]．本手法を使えば，限られた時間に効率的に動画コンテンツを詰め込むのではなく，限られた時間にユーザが見たいコンテンツを時間伸縮によってより最適に詰め込むことも可能であると考えている．消費者が大量のコンテンツとうまく向き合っていくための手法は多種多様にあり，複数の手法を組み合わせることもできる．今後，消費者及び生産者が大量の動画コンテンツと向き合うための一つの解決策となる手法の実現を目指していきたい．

謝辞　本研究の一部は，日本学術振興会特別研究員奨励費及び，JST CREST「OngaCREST プロジェクト」の支援を受けて実施された．

8. References

[1] Money, A. G. and Agius, H.: Video summarisation: A conceptual framework and survey of the state of the art, *Journal of Visual Communication and Image Representation*, Vol. 19, No. 2, pp. 121–143 (2008).
[2] 河村俊哉，福里　司，平井辰典，森島繁生：ラリーシーンに着目した映像自動要約によるラケットスポーツ動画鑑賞システム，情処学論，Vol. 56, No. 3, pp. 1028–1038 (2015).
[3] Tjondronegoro, D., Chen, Y.-P. P. and Pham, B.: Integrating highlights for more complete sports video summarization, *IEEE multimedia*, Vol. 11, No. 4, pp. 22–37 (2004).
[4] DeMenthon, D., Kobla, V. and Doermann, D.: Video Summarization by Curve Simplification, *Proc. of ACMM1998*, pp. 211–218 (1998).
[5] Smith, M. and Kanade, T.: Video skimming and characterization through the combination of image and language understanding, *Proc. of CBAIVL1998*, pp. 61–70 (1998).
[6] 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速再生方式による映像の高速鑑賞システムの検討，情処 HCI 研報，Vol. 2012, No. 13, pp. 1–5 (2012).
[7] Peker, K., Divakaran, A. and Sun, H.: Constant pace skimming and temporal sub-sampling of video using motion activity, *Proc. of ICIP2001*, Vol. 3, pp. 414–417 (2001).
[8] Cheng, K.-Y., Luo, S.-J., Chen, B.-Y. and Chu, H.-H.: SmartPlayer: user-centric video fast-forwarding, *Proc. of CHI2009*, pp. 789–798 (2009).
[9] 伊藤秀和，濱川　礼：限られた視聴時間内における動画の効果的な時間短縮手法，信学技報，Vol. 108, No. 489, pp. 23–28 (2009).
[10] 都木　徹：放送における話速変換：話者や音環境の多様性への対応，音響誌，Vol. 54, No. 7, pp. 533–538 (1998).
[11] 佐藤晴紀，平井辰典，中野倫靖，後藤真孝，森島繁生：映像の盛り上がり箇所に音楽のサビを同期させる BGM 付加支援手法，情処音楽情報科学研報，Vol. 2015, No. 10, pp. 1–6 (2015).
[12] Berthouzoz, F., Li, W. and Agrawala, M.: Tools for Placing Cuts and Transitions in Interview Video, *ACM Trans. Graph.*, Vol. 31, No. 4, pp. 67:1–67:8 (2012).
[13] Avidan, S. and Shamir, A.: Seam Carving for Content-aware Image Resizing, *ACM Trans. Graph.*, Vol. 26, No. 3 (2007).
[14] Rubinstein, M., Shamir, A. and Avidan, S.: Improved Seam Carving for Video Retargeting, *ACM Trans. Graph.*, Vol. 27, No. 3, pp. 16:1–16:9 (2008).
[15] Lartillot, O., Toiviainen, P. and Eerola, T.: A Matlab Toolbox for Music Information Retrieval, *Data Analysis, Machine Learning and Applications*, pp. 261–268 (2008).
[16] 飯塚太郎，Yue, Y.，土橋宣典，西田友是：人間の知覚特性を考慮した音と映像の特徴検出および調和の許容時間を考慮したマッチング，情処 AVM 研報，Vol. 2008, No. 124, pp. 99–104 (2008).
[17] 渡邊恵太，石川直樹，栗原一貴，稲見昌彦，五十嵐健夫：TimeFiller:生活を無理なくコンテンツで満たすメディアプラットフォーム，WISS2012 論文集，pp. 13–18 (2012).