論文の内容の要旨

論文題目 Shot-Based Adaptive Sampling Approach to Video Summarization

(適応的なサンプリングアプローチによるショットに
基づいた映像要約に関する研究)

氏名　クーハロッチャナノン　ナッグン

近年、デジタル・ビデオが広く普及し様々な領域で利用されるようになってきた．この背景にはインターネットの広帯域化，ユーザ数の増大に伴った映像を含むマルチメディア技術の発達もある． また，DVD，デジタル放送によってハイビジョンクラスの高品質な映像，音声が消費者へ提供されるようになり，デジタル・ビデオ市場は急激な成長を見せている．さらにデジタル技術の進歩によってデジカメやビデオカメラを用いた画像の撮影，映像の撮影は非常に容易なものとなった．現在では様々な企業や，大学，個人が大容量のストレージを持ち，大量のデジタル映像コンテンツ(ニュース，スポーツ，教育ビデオ，広告，ホームビデオなど)を保有している．映像技術の急速な発展により，このような大量の映像のデジタル保存が可能になった．しかしその結果，ビデオの再生，検索には非常に多大な時間が必要とされる状況が生じており，効率的な映像の取り扱いのために映像の要約や編集のための手法が強く望まれている．

本論文では，ショット内容に基づいた適応的なサンプリングアプローチについて提案する．

本論文内容は大きく２つの点にまとめられる．

１）適応的なサンプリングによる要約手法

２）ユーザのフィードバックによる 要約の改善

以下論文内容の概略を記す．まず、本論文では、映像要約を目的としたショットに基づいた適応的なサンプリングアプローチを提案する。本手法では，あるショットを代表するフレーム(R フレーム)のグループを表すことにより映像要約を実現する． 本アルゴリズムはオリジナルビデオに現れる各ショットからＲフレームを取得している．従来法では，最初のフレームまたは固定されたＮ長のフレーム群がＲフレームとして抽出されていた．しかし，本論文で提案するアルゴリズムでは，それぞれのショットの中で異なる数のフレームをＲフレームとして抽出する．具体的には，フレームの非類似度が高いショットのフレームを抽出している．

抽出するRフレームの数はショットの長さとショットの動き情報により決定する．
ショット境界はシティブロックアルゴリズムを用いて隣接する２つのフレーム間の距離を求めることによって決定する．シティブロックアルゴリズムによって求めた距離が閾値より大きければショット境界とする．ショット内の動き情報は，MPEG ビデオであればMPEGの動き情報を利用し，たとえばAVIビデオであればブロックマッチングアルゴリズムによって求める．通常，ショット長が長いほどその区間に含まれる情報量は多くなるので，その分要約映像に含まれるキーフレームの数が多い方が自然である．しかし，ショット長が長いにもかかわらず含まれる動き情報が少ないショットから得られるキーフレームは類似した変化の乏しいフレームを多く抽出してしまう． 言い換えれば，全体の平均よりも多い動き情報を持つショットはより変化の激しいシーンを含んでいると考えられる．従って，変化の激しいシーンからより多くのキーフレームが抽出されるべきである．このように抽出されたRフレームからスムースパラメータを用いて要約映像を滑らかに再構成する．
実験において，ホームビデオを異なる長さに要約を行った．評価はその要約圧縮率とユーザによる明快性，簡潔性，一致性の３つの観点から行った．明快性は映像内容の理解のしやすさ，簡潔性は要約映像の中に元の映像の情報がどれほど反映されているかを表し，一致性は要約率の変化による変化の少なさを表す．簡潔性は圧縮率が低ければ高くなり，明快性は低くなる．また，映像中にイベントがほとんど含まれていなければ一致性は圧縮率に関係なくほぼ一定である．ここでいうイベントとは映像に含まれる場所の変化による画像全体の変化のことを言う．たとえば，屋外から地下鉄に乗りまた屋外にでるといった映像のほうが，同じ部屋の中での映像よりもイベントが豊富であると定義する．このことにより，一致性の値はイベントが少なければ，圧縮率が高い場合でも低い場合でも要約映像に含まれる画像の変化が乏しいためほぼ一定の値となる．逆にいえば，イベントが豊富な映像では圧縮率が高いときと低いときには大きな差が生じることになる．

要約された映像の内容では，できるだけ多くのショットから情報を得るための制御が必要である．本アルゴリズムではユーザがショット数を制御可能であり，同一の要約時間であっても，ショット数が多い場合，ショット数が少なくRフレームが多い場合を制御できる．

比較実験として，均一的なサンプリング手法及び適応的なサンプリング手法であるRシーケンスと本手法との比較を行った．実験よりショットに基づく本手法の方が，適切にショットからサンプリングされていることが確認できた．

次に論文の後半であるユーザのフィードバックに関して述べる．通常，作成された要約映像はある条件を満たした最適な映像であるが，本手法ではユーザに結果のフィードバックを可能とすることで，よりユーザの望む要約映像の作成を可能にする．本手法では，サポートベクタマシン（SVM）を利用することでその目的を実現している．従来の検索フィードバックはユーザが好みの重みを手動で入力する必要があるが，適切な値の入力

は非常に困難である．従って特別な好みの重みを指定するよりも，好きであるか，嫌いであるかを与える方が適切であろうと考えられる．提案手法では，ユーザはポジティブなフィードバックだけでなく，ネガティブなフィードバックも与えることが出来る．適切な画像に対する新しい好みに関する重みは SVM によって与えられる．

本アルゴリズムでは、まず抽出された R フレームのグループを正または負と分類され，これらの結果はトレーニングデータとして使用される．SVM はオリジナル映像を 2 クラス(ポジティブとネガティブ)に分類するために用いる．分類されたデータはフレームと超平面の間の距離も割り当てられる. 分類されたデータからフレームを検索するために, 本手法では超平面と検索クエリーとの距離, 超平面とポジティブなデータを比較している. 検索結果のフレームは超平面と検索クエリーとの距離との差が最も小さいポジティブデータが選ばれる. 検索されたフレームはユーザに提示され, ユーザが結果に満足しなければ, 再びユーザが満足する結果を得るまでフィードバックを繰り返す．最終的に、検索されたポジティブなフレームのグループは平滑化され要約映像にまとめられる．

ＳＶＭにおける実験で，１）高レベル特徴量の類似性（ユーザの好み）２）低レベル特徴量の類似性（色の類似性）の２つのタイプのトレーニングデータが生成される．２つのトレーニングデータを用いた分類結果では検索されたフレームはポジティブな例と類似していた．また，ネガティブなフレームに類似するフレームは結果には含まれなかった．

次に，ＳＶＭにおける多項式カーネルと RBF (Radial based function) kernel カーネルとの比較を行う．結果からどちらのカーネルでも同様なポジティブな例を検索することがわかった．しかしながら，RBF カーネルを用いて得られた検索フレームは多項式カーネルから得られた検索フレームより類似度が高い傾向にある．

本論文では，ショットに基づいた適応的なサンプリングアプローチによる映像要約手法を提案した．要約映像は R フレームの組から構成される．各ショットに含まれる R フレームの数はショット長とショット内の動き情報から求められ，要約映像に取り入れられる R フレームは適応的なサンプリングアルゴリズムによって決められる． 要約結果の評価はユーザによって行った．従来のサンプリング手法との比較により，本手法のほうが従来の手法より良好な結果が得られた．また，検索フィードバックを用いた映像要約手法についても提案した．本手法では，学習にサポートベクタマシンを用いた．実験では，２つのトレーニングカーネルについて異なるパラメータで比較，評価を行い，良好な結果を得ることが確認できた．