

対話型進化計算によるユーザの嗜好を反映した ダイジェスト動画の生成

中村 宏^{1,a)} 小野 智司^{1,b)} 中山 茂¹

概要: 本研究では対話的な進化計算を用いて利用者の嗜好を反映した映像のダイジェスト生成手法を提案する。ダイジェストの適切さは質的評価と量的評価の2つの指標から評価する。質的評価はユーザの嗜好の反映度である。量的評価は再生時間の適切さと冗長性の少なさである。冗長性の評価やユーザ評価の推定に、Bag of Visual Wordsを使用した。提案手法ではユーザ評価の推定に特定のジャンルの映像に依存した特徴量を用いないため、ホームビデオ等の映像にも適用可能である。

キーワード: ビデオダイジェスト, 対話型進化計算, Bag of Visual Words

Generation of video digest that reflects User Preference by Interactive Evolutionary Computation

Abstract: This paper proposes a video digest generation method based on interactive evolutionary computation. The proposed method evaluates solution candidates from the viewpoints of qualitative and quantitative criteria. As a qualitative evaluation, the proposed method maximizes user preference, and as quantitative evaluation, the method minimizes the redundancy and the inappropriateness of video length. The method uses features based on Bag of Visual Words, which are not depend on video categories and users, enabling even to summarize home videos.

Keywords: Video digest, Interactive Evolutionary Computation, Bag of Visual Words

1. はじめに

個人向けの撮影機器の低価格化や記録媒体の大容量化、インターネット動画共有サービスの登場によって、動画から重要な要素を抽出して編集する作業の必要性が高まっている。様々なソフトウェアが販売、公開されているが、再生時間の長い動画や多数の動画を編集する作業は多くの労力を必要とする。また、重要だと感じる要素は人によって個人差がある場合があり、画一的な基準では意図した部分を十分に抽出できない。

本研究では、対話型進化計算を用いてユーザの嗜好を反

映した動画要約手法を提案する。進化計算により生成されたダイジェストをユーザが評価することで、ユーザに最適なダイジェストを生成する。本手法におけるダイジェストの適応度は、ユーザの嗜好に基づく質的評価と映像の解析結果に基づく量的評価からなる。また、本提案手法では動画のジャンル固有のオブジェクトの検出といった特徴量を用いないため、ホームビデオなど、特徴からの意味解釈が困難である映像に対しても適用可能である。

2. 関連研究

益満らは、ユーザの視聴行動から要約映像を生成する手法を提案した [1]。この研究では、ユーザが視聴した映像と類似したフレームは重要度が高く、ユーザが早送りした映像と類似したフレームは重要度が低いものとして重要度を推定する。要約映像の生成にはコンテンツやユーザに依存する特徴は用いず、推定された重要度及び、画像の類似度のみを用いる。

¹ 鹿児島大学理工学研究科情報生体システム工学専攻
Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University
Kagoshima, Kagoshima, 890-0065, Japan

a) k8088145@kadai.jp

b) ono@ibe.kagoshima-u.ac.jp

Parshin らは、動画の取捨選択を行うフィルタをユーザに定義させて要約映像を生成する手法を提案した [2]。この研究では、映像をショット単位に分割し、映像の特徴を分析する。ユーザは、動画の特徴量に従って選別を行うフィルタの設定を変更することでユーザの嗜好に合った要約映像が生成される。

3. 提案手法

3.1 基本方針

益満らの手法 [1] は、ユーザの視聴行動と注視で要約映像が生成されるため、視聴者の好むシーンに集中した要約が生成される。Parshin らの手法 [2] は、個々の特徴量と選択されるショットの関連性について理解しなければ意図したシーンを選択するのは難しい。

本研究では、対話的進化計算によってユーザの嗜好を取り入れつつ、ユーザの意図とは異なるダイジェストも含めて進化計算によって複数の個体を生成することでユーザの発散的思考を支援する手法を提案する。また、ダイジェスト生成に必要なユーザの操作をダイジェストの評価のみにすることで、ユーザの負担を軽減する。

3.2 特徴量

提案手法ではダイジェストの評価に用いるため Bag of Visual Words (以下, BoVW) を、ダイジェスト対象の動画を分割した、クリップを表わす特徴として用いる。クリップ内の全フレームの BoVW を求め、その平均値を求める。その平均値に最も類似したフレームをクリップのキーフレームとし、キーフレームの BoVW をクリップを表わす特徴ベクトルとする。

提案手法では動画の 1 フレームを 1,000 次元の特徴ベクトルで表現する。Visual Words の学習画像としてダイジェスト対象動画から 1 秒間隔で切り出したフレームを用いる。また、位置情報を導入するため入力画像を 4×4 のグリッド状に分割し、RGB それぞれのチャンネルに対して SURF 特徴量を抽出し、1 つのベクトルとして結合する方式をとる。

3.3 遺伝的アルゴリズムの適用

提案手法では、進化計算に遺伝的アルゴリズム (Genetic Algorithm 以下, GA) を用いる。GA に適用するためにダイジェストの対象となる動画を N 個のクリップに分割し、それぞれのクリップに 0 から $N - 1$ までの番号を割り振る。この割り振った番号を遺伝子とする。また、クリップの再生時間はそれぞれ異なるため、ダイジェストのクリップ数は生成されたダイジェストごとに異なる。従って、提案手法では遺伝子長が可変長の GA を用いる。

本提案手法における個体 x の適応度 $F(x)$ は、質的評価 $P(x)$ と量的評価 $Q(x)$ の重み付き線形和で表す。

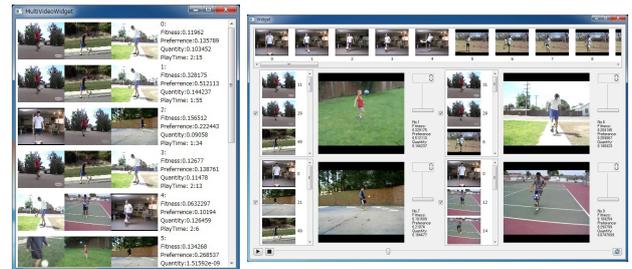


図 1 提案手法のインターフェース

$$F(x) = w_p \times P(x) + (1.0 - w_p) \times Q(x) \quad (1)$$

w_p は質的評価の重みであり、0 以上 1.0 以下の値を示す。

質的評価 $P(x)$ は、ダイジェストに対するユーザの嗜好の度合いであり、図 1 に示すインターフェースを通じてユーザが指定する。ユーザが評価していない個体は最も類似しているユーザの評価済み個体 s の類似度 $S(x, s)$ とその正規化済み評価点 p_s から推定する。

$$P(x) = p_s \times S(x, s) \quad (2)$$

類似度 $S(x, s)$ は、クリップ間の類似度の平均値とする。

量的評価 $Q(x)$ は再生時間の適切さ $D(x)$ と非冗長性 $R(x)$ の積とする。

$$Q(x) = D(x) \times R(x) \quad (3)$$

$D(x)$ は、ユーザの設定した再生時間 t_d に対する個体 x の再生時間 t_x の一致度、 $R(x)$ は個体 x 中のクリップ間の類似度の平均値とする。

$$D(x) = \exp \left[-\frac{\ln 2 \times (2(t_x - t_d)^2)}{t_d^2} \right]^2 \quad (4)$$

$$R(x) = \frac{\sum_{i=0}^{N_x-1} \sum_{j=i+1}^{N_x-1} Sim(x_i, x_j)}{\sum_{k=1}^{N_x-1} k} \quad (5)$$

N_x は個体 x のクリップの総数、 x_i, x_j は個体 x の先頭から i 番目、 j 番目のクリップの番号とする。

4. おわりに

本稿では、対話的進化計算を用いることでユーザの嗜好を反映した動画要約を行う手法を提案した。提案手法は、コンテンツに依存しない特徴量を用いる。また、ユーザへの負担を軽減し、ユーザの発散的思考を支援する。

今後、被験者実験を行い本手法の有効性の検証する。

参考文献

- [1] 健 益満, 富夫越後: 映像重要度を用いたパーソナライズ要約映像作成手法, 信学論. D-II, 情報・システム, II-パターン処理, Vol. 84, No. 8, pp. 1848-1855 (2001).
- [2] Parshin, V. and Chen, L.: Video Summarization Based on User-Defined Constraints and Preferences, *Computer-Assisted Information Retrieval, 7th International Conference, 2004.*, pp. 18-24 (2004).