

フレーム間差分に着目した イラストレーションメイキング動画のキーフレーム要約

高橋 リサ†

川戸 美輝†

藤波香織††

† 東京農工大学院 工学府 情報工学専攻

†† 東京農工大学 工学研究院 先端情報科学部門

1 はじめに

インターネットの普及により動画投稿サイトではあらゆる動画が投稿され、イラストレーション等の創作活動の制作過程を動画として発信する「メイキング動画」の投稿も一般的になりつつある。これは初学者にとって優れた学習教材である一方で全過程を視聴するために多くの時間を要し、ユーザが自身の視聴目的に沿った映像を探索することが困難である。この問題は多数のジャンルの動画視聴時に生じており、その対策として動画の要約に関する研究が広く行われている。動画要約の研究において用いられる手法として、要点を保持したまま映像を短縮化する動的要約 [1], 映像から特徴的な場面をキーフレームとして抜粋する静的要約 [2][3] などが存在する。しかし、これらの要約手法の多くは対象をフレーム情報が大きく変化する実世界の動画やアニメーションに限定し、フレーム情報が緩やかに変化する創作活動の制作過程動画において適切に要約を行うことは困難である。したがって、本研究ではイラストレーションのメイキング動画に対応した新たな要約システムを提案し、その設計及び開発を行う。評価実験では既存手法におけるキーフレーム生成方法との性能比較及びメイキング動画の静的要約に適した特徴量の評価を実施する。

2 提案システム

メイキング動画は人によって視聴目的が異なるため、冗長な情報を削減し動画の時間短縮を図る動的要約は適さない。したがって、本システムは重要度を定義付けし要約映像を生成する動的要約手法ではなく、キーフレームと動画の再生プレーヤーを組み合わせた静的要約手法を採用する。本システムは、図 1 に示すように、動画再生プレーヤー下部に生成したキーフレームの一覧を表示し、キーフレームが選択された際に表示しているキーフレームの時間までメイキング動画を遷移させることで、ユーザが見たい項目・内容を検索することを支援する。例としてユーザが肌の着色工程に興味がある場合、ユーザが提案システム上でキーフレームから肌の着色が実施されていると思われる部分を選択すると、動画をキーフレームが表示されている時間まで遷移させる。本システムは図 2 に示す通り特徴量生成部、キーフレーム生成部、動画再生部からなる。特徴量生成部において生成した特徴量を使用してキーフレーム生成部にてクラスタリングを実行し、各クラスターの先頭時間のフレームをキーフレームとして動画プレーヤー下部に表示する。

2.1 特徴量生成部

メイキング動画は少しずつ描かれていく過程でペン先が移動しキャンバスに色が塗られていくため、ペン先を



図 1 動画再生部

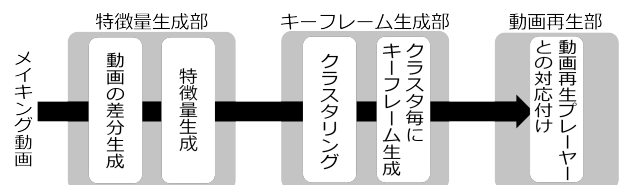


図 2 システム概要

移動物体と捉え移動物体検出方法の 1 つであるフレーム差分法を利用することで、新たに描画された部分を抽出することが可能であると考えた。この描画部分のヒストグラムを特徴量として利用し描画色に基づくフレームの分類を試みる。ヒストグラムの bin 数は Avila らの研究結果 [2] を元に本システムでは色空間 1 つにつき 16 とする。また、動画の時間的連続性に対応するために動画の先頭から経過した時間も特徴量として併せて採用する。

2.2 キーフレーム生成部

視聴者が容易に探索可能なキーフレームを生成するためにはキーフレームが内容ごとに分割されている必要がある。したがって、類似情報を 1 つのクラスターとして分類する手法であるクラスタリングを使用し、キーフレームを生成する。初期グループの決定方法は Avila らの改良型 k-means 法 [2] を参考に、時間的先頭からグループ化を行う。本研究では実験時に人物イラストのメイキング動画を使用するため、その工程は「ラフ + 線画」「肌の色塗り」「髪の色塗り」「服の色塗り」「小物の色塗り」の 5 項目に大別できると考え、k の値は 5 とする。また、同一クラスター内の要素間の時間的連続性を保持するために、クラスター内で時間的に離れた要素の集合が存在する場合には異なるクラスターとする分割処理を実施する。

2.3 動画再生部

動画再生部は実際にユーザが使用する部分であり、動画プレーヤー、キーフレーム群からなる。キーフレームは 2.2 節のクラスタリング結果を元に表示する。本システムは選択されたキーフレームが属するクラスター部分へ

An illustration making video summarization based on color difference between frames

†Lisa TAKAHASHI †Miki KAWATO ††Kaori FUJINAMI

†, ††Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology

動画を遷移及び再生することで視聴者の検索を支援するシステムであるため、表示するキーフレームの動画時間と実際に遷移する動画時間が異なると操作時にユーザの混乱を招く恐れがある。したがって、キーフレームを生成する時間は各クラスターの先頭時間とした。

3 評価実験と考察

13名の被験者を対象として色空間 (HSV) 特徴量を変数としたクラスタリング精度評価実験および既存手法における利用手法との比較評価実験を実施した。

3.1 特徴量を変数としたクラスタリング精度評価

イラストレーションのメイキング動画の静的要約に適した特徴量についての調査を目的として実施した。特徴量の候補は時間特徴量に加えて、色相、彩度、明度に関するヒストグラム及びこれら三つ全てを結合したヒストグラムの四通りである。二本のメイキング動画を使用し、被験者に動画ごとに各手法によるキーフレームの質を五段階リッカート尺度で回答を要求し、“非常に良い”を5点、“非常に悪い”を1点としてスコア化した。スコア結果を図3に示す。各特徴量によるキーフレーム要約のスコア結果を群として、有意水準5%でSteel-Dwass法による多重比較検定を行なった。その結果、動画1においては全ての特徴量間で有意差が認められず、動画2においては、色相-明度間と色相-色相彩度明度間で有意差が認められた。また、色相-彩度間は多重比較検定では有意的な差は見られなかったものの、スコアの平均値は色相が僅かに上回っていた。本結果を元に次節の実験では「時間特徴量、描画差分の色相に関するヒストグラム(16bin)」で構成される17次元の特徴量をクラスタリング時の特徴量として使用した。

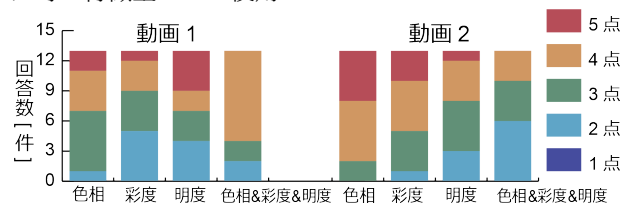


図3 特徴量ごとのキーフレームの質に対する評点

3.2 既存研究での利用手法との比較評価

提案手法の特徴量生成方法と通常の動画要約を目的とする既存手法の特徴量生成方法を比較することで提案システムの有効性を評価した。比較する既存手法はAvilaらのVSUMM[2]である。この手法は動画のフレームを特徴量としてクラスタリングを実施することでキーフレームを生成しており、本システムのキーフレーム生成手法と類似しているため選出した。VSUMMでは動画のフレーム全体の色相空間のヒストグラムをk-means法でクラスタリングすることで、キーフレームを導出している。四本のメイキング動画を使用し、被験者に動画ごとに各手法によるキーフレームの質および提示するキーフレームの個数を五段階リッカート尺度で回答を要求し、スコア化した。なお、単盲検法により被験者は適用された手法を知らない。スコア結果を図4に示す。本手法及び既存手法のスコアに対して有意水準を5%としてWilcoxon順位和検定を実施した。その結果、キーフレームの個数スコアにおいては、動画1と4において提案手法が有意にスコアが高くなったが、動画3においては既存手法が有意にスコアが高い結果となった。キー

フレームの質のスコアは全ての動画で有意差が認められなかった。

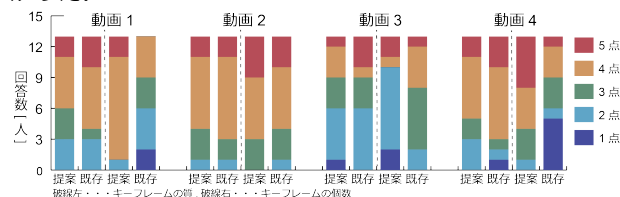


図4 要約手法ごとのキーフレームの質および個数に対する評点 (破線左・・・キーフレームの質, 破線右・・・キーフレームの個数)

3.3 考察

3.1節の実験では一部の動画においてのみ色相特徴量を使用したキーフレームが有意にスコアが高い結果となった。この原因としてツール領域・回転拡大操作の有無が挙げられる。有意差が見られなかった動画はワークスペース全体が表示される動画であり、キャンパスの回転や拡大も存在したが、有意差が見られた動画はキャンパスのみを固定キャプチャしたメイキング動画であったため、これらの有無がフレーム差分生成時に影響したと予測される。

また、3.2節の実験ではその結果からキーフレームの個数は使用する動画によって既存手法と比較して有意にスコアが高い動画、低い動画が存在することが分かる。使用動画の完成時のイラストのHSV空間それぞれのヒストグラムの分散を分析した結果、動画1及び4は色相の分散が最も大きく、動画2は彩度、動画3は明度の分散が最も大きいことが判明した。したがって、色空間の分散は提案手法によるキーフレーム生成個数に影響することが推測され、色相分散が大きい動画の場合にはよりユーザに好まれる個数のキーフレームを提示する可能性が示唆される。また、キーフレームの質のスコアは全ての動画で有意差が見られなかった。各評価実験内で被験者がスコア付けた理由を参照すると、工程やその粒度など視聴を希望する内容が被験者によって多様であり、この個人のメイキング動画の視聴目的の差異が影響したと推測される。

4 おわりに

本論文ではイラストレーションのメイキング動画に対応した要約システムの設計及び開発を行い、既存手法におけるキーフレーム生成方法との性能比較及びメイキング動画の静的要約に適した特徴量の評価を実施した。実験の結果から、色相分散が強い動画においては本手法は既存手法と比較してキーフレームの個数の観点から良い評価が得られた。更に、各動画の色空間の分散やツール領域・回転拡大操作の有無がこの評価に影響を及ぼす可能性があるため、使用する動画に応じた特徴量の設定や例外処理等が有効であることが示唆された。

参考文献

- [1] D. Zhong, S.-F. Chang. Structure analysis of sports video using domain models. In *Proc. of IEEE ICME 2001*, pp. 713-716.
- [2] S. E. F. De Avila *et al.* VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recognition Letters*, Vol. 32(1), pp. 56-68, 2011.
- [3] S. Uchihashi *et al.* Video manga: generating semantically meaningful video summaries. In *Proc. of ACM Multimedia 1999*, pp. 383-392.