

# 動画からの感情抽出および 感情遷移によるストーリーの自動構築

浅田 裕也 富士本 達矢 濱川 礼  
中京大学 情報科学部 情報科学科

## 1.はじめに

本研究は、ストーリーを持った動画から感情を抽出し、感情の遷移によって異なるストーリーを構築するシステムの開発を目的としている。

近年、動画投稿サイト(Youtube<sup>[1]</sup>やニコニコ動画(RC2)<sup>[2]</sup>)などでユニークな創作動画を多人数で共有し、コメントなどによって評価を行うなど、動画によるコミュニケーションが活発化し、利用者が急増している。ニコニコ動画(RC2)では、動画の評価方法のひとつに再生数によるランキング形式が取られており、話題性のある面白い動画の再生数は数 100 万に及ぶこともある。しかし、このような動画とセンスや動画編集技術の乏しい動画との差は如実に現れ、投稿した動画が必ずしも高評価を得るわけではない。このようなサイトに動画を投稿する際には、面白い動画を創作する豊かなセンスと高い動画編集技術が求められる。そこで、センスが乏しい、又は動画編集技術を持っていない利用者のために動画編集作業をコンピュータの手に委ねることで、作成者への負担を大幅に軽減し、またセンスに関係なく面白い動画を生成できないかと考えた。

我々は、動画に対して画像解析を行ってシーン毎に分割し、分割されたシーンに対して音声解析を行って感情を抽出する。そして、ストーリー性を持った感情遷移を定義した結合規則に沿ってシーンを並び替えることで異なるストーリーを構築するシステムの開発を行った。

## 2.本システムの特徴

動画を視聴する際、ストーリーの進行に伴って視聴者が受ける感情は変化することから、感情の遷移を利用してストーリーの構築を自動化する。

本システムと関連研究のWindowsムービーメーカー<sup>[3]</sup>、Highlight Creator<sup>[4]</sup>との比較を表 1 にまとめた。他研究と比べて異なるストーリーを生成することと、音声から特徴量として感情を抽出することに特化している。

表 1.本システムと関連研究との比較

	目的	動画編集	自動化手法
Windowsムービーメーカー	動画編集	手動	なし
Highlight Creator	要約	自動	映像+音声の特徴解析 映像：動きの大きさ 音声：歓声の大きさ
本システム	異なるストーリーの生成	自動	映像+音声の特徴解析 映像：色変化の大きさ 音声：感情

## 3.結合規則

ストーリーを構築するために 6 種の感情<sup>[5]</sup>(幸福、驚き、怒り、嫌悪、恐れ、悲しみ)の比率を用いて、感情の遷移を結合規則として定義した。これに沿ってシーンを並べることでストーリー性を持つ動画の生成を実現した。

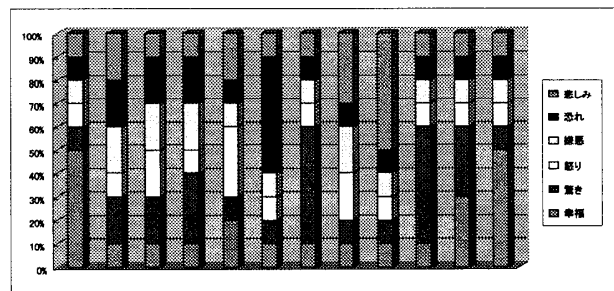


図 1.結合規則

図 1 はハッピーエンドの起承転結構造を作るため、シーン数 12(起・承・転・結それぞれの核となる感情比率とその前後への変化過程を表す 2 つの感情比率)で表現した例である。悲しみの感情が強く現れた後、幸福の感情が強く現れている。反対にバッドエンドでは幸福の感情が強く現れた後、悲しみの感情が強く現れる。ストーリーとして最も重要な転・結において、感情の起伏を大きく持たせることでストーリー性を強く持たせた。

## 4.システムの全体構成

本システムは、『分割処理』、『感情抽出処理』、『結合処理』の 3 つの基本機能から構成されている。

図 2 にシステムの構成図を示す。

The feelings extraction from an animation and the automatic construct system of the story by the feelings transition  
Yuya Asada, Tatsuya Fujimoto, and Rei Hamakawa  
Chukyo University Department of information science

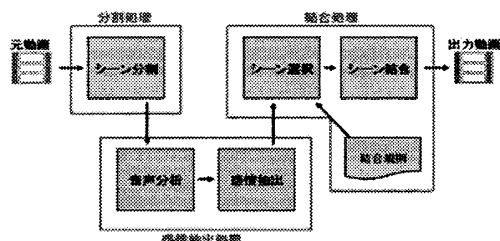


図 2.システム構成図

#### 4.1.分割処理

入力された動画をシーン毎に分割する。画像情報を調べ、画像が大きく切替わるカット点を分割カイ二乗検定法を用いて検出し、分割する。

#### 4.2.感情抽出処理

分割されたシーンの音声データに離散フーリエ変換処理を行い、振幅スペクトルを生成する。音声に含まれる感情の特徴が周波数と音圧に現れることから、6種の感情（幸福、驚き、怒り、嫌悪、恐れ、悲しみ）に対応した分析区間を図3のように振幅スペクトル上に設置する。この分析区間を感情窓として定義した。感情窓内の離散点の出現回数により、各感情を数値で表した感情情報を生成し、シーンに付与する。

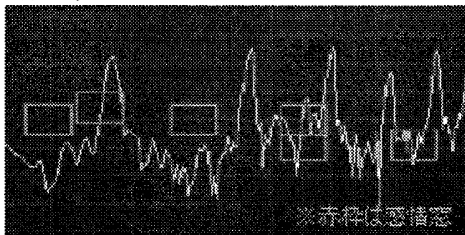


図 3.振幅スペクトルと感情窓

#### 4.3.結合処理

各シーンに付与された感情情報を、各感情の割合で表した感情比率に変換する。変換した感情情報と結合規則に定義した感情比率を比較し、各感情の差を求める。差が閾値以下の感情の数を評価値とし、全シーンに対して同様の処理を行う。評価値が最も大きいシーンを選択、結合し、動画を出力する。図4.に評価値算出の例を示す

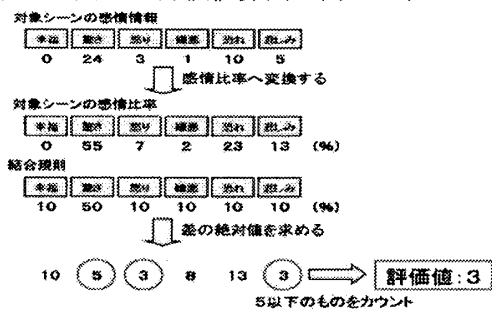


図 4.評価値算出例

## 5.評価

動画ファイルの入力から完成した動画の出力までの処理時間を計測した結果を表2に示す。

表 2.処理に要した時間

	計測時間
10分の動画	12分58秒
30分の動画	37分44秒

今回計測に使用したマシンは、CPU:Pentium4 2.8GHz, メモリ:1GB, OS:Windows XP SP2である。

処理時間の大半を動画処理が占めており、高速な動画処理を行う必要がある。

本システムによってテレビアニメ、テレビドラマそれぞれから出力された動画を用いて評価を行った。本研究室に在籍する学生10名に出力動画がストーリーを持っているか、5段階評価でアンケートを行った結果、平均2.9となった。この結果から、ストーリーを持たない動画が出力されないよう、シーン選択アルゴリズムを再検討する必要がある。

## 6.まとめと今後の課題

本研究では、動画からの感情抽出を行い、感情遷移を定義した結合規則によってストーリーを自動構築するシステムを開発した。

音声処理において、音圧によって感情情報が取得できない場合がある。音圧に左右されない音声解析手法をとることで改善する。

現状では、音声解析のみによるストーリー性の実現であるため、映像ではストーリーが成立しないことがある。音声解析だけでなく画像解析も取り入れることで音声、映像双方においてストーリーを持つ動画を生成することを目的とする。

## 参考文献

- [1] Youtube <http://jp.youtube.com/>
- [2] ニコニコ動画(RC2) <http://www.nicovideo.jp/>
- [3] Windowsムービーメーカー <http://www.microsoft.com/japan/windowsxp/using/moviemaker/default.msp>
- [4] Highlight Creator <http://mmm.kddilabs.jp/ja/hc/>
- [5] Ekman, P. & Friesen, W. V., Unmasking the face. A guide to recognizing emotions from facial clues, 1975
- [6] 雨宮 俊彦・水谷 聡秀, 「はいそうですね—音声の印象の分析のころみ—」, 関西大学