

テレビ番組の字幕画面抽出

Extraction of caption images in TV program

石綿 翔平

高橋 正信

D-11

Shohei Ishiwata

Masanobu Takahashi

芝浦工業大学

システム理工学部

Shibaura Institute of Technology

Collage of Systems Engineering Science

1. 背景

多くのテレビ番組が地上波で放送されているが、興味のある番組を視たくても時間が無く断念することがある。短時間で番組の内容の情報を把握することができれば、視聴者にとって有益である。そこで、番組内容を短時間で把握するために、「字幕画面のみ」を利用者が視られるようにする機能を実現する。字幕画面を視ることで画像から雰囲気や、セリフから内容を短時間で把握することができる。

字幕のついたテレビ番組が対象となるが、関東圏の主要なテレビ局の字幕放送時間の割合[1]は、低い日本テレビ放送網でも 44.8%、最も高い NHK (総合)では 72.3%あり、このシステムが実現すれば多くの番組で利用できる。

2. 目的

字幕のあるドラマやアニメなどのテレビ番組の字幕画面を抽出し、字幕画面のみを表示して見られるようにすることにより、番組内容をより短い時間で把握できるようにする。処理を極力自動化するとともに、同じ字幕の画面を1枚だけとすることで字幕画面の枚数を削減する。

3. 手法

本研究ではカメラの位置合わせ以外の動作をできる限り自動化する。BD 再生ソフトの操作についても WindowsAPI を用いて自動化した。以下に処理の流れを示す。

3.1 撮影

カメラの位置を手動で設定後、自動で字幕ありの映像と字幕なしの映像から同じタイミングで撮影する。両者の同期には動きのあるフレームを用いる。フレーム間差分が大きいフレームを1枚決定し、そこから字幕なしの画面を一定間隔(今回は 0.5 秒)で撮影する。続いてコマ送りで同じフレームを検出し、そこから字幕ありで一定間隔で撮影する。

3.2 撮影画像から画面画像を抽出

撮影画像を読み込み、座標を指定して映像が映っている領域(画面画像)を切り取る。

3.3 画面画像から字幕の有無両画像の差分を抽出

同じタイミングの字幕あり画像となし画像の差分を二値化し(図1)、オープニングとクロージング処理でノイズを減らして字幕の有無を判別しやすくする。

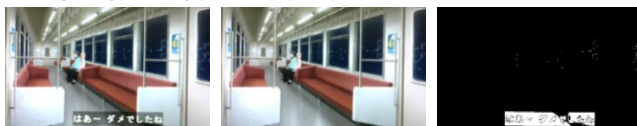


図1 字幕ありとなしの画面画像と差分の二値化画像

3.4 字幕の有無の判別

差分画像中の差分が小さければ字幕がないと判定。差

分が大きければ字幕があると判定(図 2)。字幕のある画像だけを抽出する。

3.5 グループ化と代表選択

字幕のある二値化画像の前後の差分を取得し、差分が小さいものを同一グループと判定。同一グループの真ん中の画像を1枚選択し、それに対応する画面画像(字幕あり)を代表画像として抽出する。

3.6 表示

代表画像の集合をスライドショーで表示することによりユーザーが番組内容を短時間で把握できるようにする。

4. 実験

実験にはテレビアニメ(やはり俺の青春ラブコメは間違っている。続)の第 10 話[2]を録画した映像を用いた。再生ソフトには CyberLink PowerDVD 10 を、カメラはウェブカメラ (Logicool HD Webcam C310)を用いた。なお、PowerDVD の動作は WindowsAPI を用いたショートカットキー入力により自動化した。撮影画像の総枚数は 3360 枚である。

字幕の種類は 292 種類であったが、本手法により自動抽出された画像は 865 枚であった。その中で正しく抽出された字幕画面は 240 枚であった。再現率(抽出された種類数/字幕の種類数)は 82.2%、適合率(抽出された種類数/字幕画面として抽出された総数)は 27.7%であった。また、誤抽出された画像は 625 枚であった。

字幕画面の 82%を抽出できたが、誤抽出が多くその低減が必要である。主な原因は撮影タイミングのズレである。今回利用した PowerDVD ではコマ送りし、巻き戻した際に全く同じコマを再現できなかった。再生開始位置にわずかもズレがあると、字幕がないにも関わらず図 3 左のように差分が大きくなり誤抽出される。また、背景と字幕の色が近い場合などで抽出された領域がオープニング処理によりさらに小さくなり、図 3 右のように字幕があるのにもかかわらず抽出できない場合もある。同期方法などを再検討し、これらの問題を解決したい。

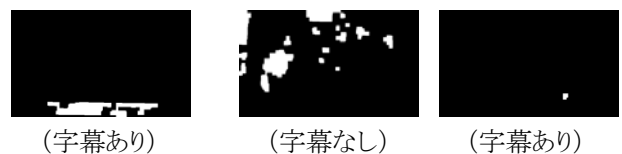


図 2 成功例

図 3 失敗例

【参考文献】

- [1] 総務省, “平成 25 年度の字幕放送等の実績”, http://www.soumu.go.jp/menu_news/s-news/01ryutsu09_02000106.html, 参照 June 22, 2015.
- [2] TBS, “やはり俺の青春ラブコメは間違っている。続”, 第 10 話, 2015 年 6 月 5 日放送.