

# テレビ番組の字幕画面抽出 —精度改善—

## Extraction of caption images in TV program —Accuracy improvement—

萩原 良耶

高橋 正信

D-11

Ryoya Hagiwara

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、視たくても時間が無く断念することがある。視るよりも時間をかけずに番組内容を把握できれば、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の実現を考えた。字幕画面を視ることで短い時間でセリフを把握することができ、画像から雰囲気も把握することもできる。

これまでの研究[1]で字幕画面抽出機能を実現しているが、カメラで撮影を行う点や、撮影タイミングの同期ズレの問題などにより、再現率(正しく抽出された字幕画面数/正しい字幕画面数)82.2%、適合率(正しく抽出された字幕画面数/抽出された画面数)27.7%と、誤抽出が多かった。

### 2. 目的

本研究では、字幕画面抽出の精度改善を目的とする。撮影方法と同期方法、および画像処理手法の改善を行う。

### 3. 手法

#### 3.1 撮影

字幕画面は、字幕ありと字幕なしの両画面を同じタイミングで撮影し、その差分を利用して抽出する。従来[1]はカメラ撮影で取得していたが、スナップショットに変更した。DVDの再生画面をスナップショットすることは違法でないことが確認できたため、再生プレーヤー(PowerDVD)で再生した画面をスナップショットする。撮影方法の変更により、カメラに関する操作を省くことができ、システムの自動化と時間短縮が実現できた。

また、撮影間隔を 0.5 秒から 1 秒に変更し、時間計測にはマイクロ秒まで計測可能な QueryPerformanceCounter 関数を採用した。同期方法の変更により、字幕領域以外に生じる不要な差分が低減した。

#### 3.2 字幕画面の抽出

従来は、字幕有無両画面の差分画像をグレースケール化、2 値化し、オープニング・クロージング処理により字幕領域を抽出したが、同期ズレの影響を軽減するために 2 値化閾値を大きくする必要があり、字幕領域が一部しか抽出されない問題があった。本研究では撮影方法が改善されたことで 2 値化閾値を小さく設定して字幕領域全体を抽出できるようになった(図 1 中央)。

次に、字幕領域を再度 2 値化して字幕領域内の字幕文字を抽出する。実験に使用した番組[2]の字幕文字は 4 色(白, 黄, 水色, 緑)あったが、このうち緑を除く 3 色についてはグレースケール化した画像を閾値 127 で 2 値化するこ

とで字幕文字のみを抽出できた。緑色の字幕はグレースケール画像の 2 値化では抽出できなかつたが、色度が他と異なる特定(緑色)の領域内にあることがわかったため、色度によって抽出した。なお、字幕が緑か他の色かの判定には字幕領域中の緑色領域の比率(字幕領域中の色度図の緑色部分に存在する面積/字幕領域の面積)を用い、比率が 5%以上の場合を緑色と判定した。なお、字幕文字の面積の割合(字幕文字の面積/字幕領域の面積)が 20%以上のものは誤抽出である可能性が高いため、全ての色についての割合が 20%未満のもののみを字幕文字および字幕画面として抽出した。図 1 右は、緑色字幕文字の抽出結果である。

次に、連続する同じ字幕の字幕画面をグループ化し、そのうち1つを最終的な字幕画面とする。従来は字幕領域を比較して差分が小さい場合を同じグループとしていたが、本研究では字幕文字を抽出できているため、字幕文字を画像として比較し、差分が小さい場合に同じグループとした。従来は字幕領域が同じであれば字幕文字が異なっても誤ってグループ化される問題があったが、字幕文字を直接比較することでこの問題を解決できた。



図 1 文字抽出(左:原画像, 中央:字幕領域, 右:字幕文字)

### 4. 実験

実験には、テレビアニメ(BANANA FISH)の第 1 話[2]を録画した映像(CM を含む 30 分)を用いた。実験の結果、全ての字幕画面 413 枚を正しく抽出することに成功した(再現率 100%)。一方、誤抽出が 108 枚存在した。その内訳は、「本編の同じ字幕を抽出」が 50 枚、「字幕がない画面を抽出」が 58 枚(そのうち CM 中が 55 枚)であった。適合率は 79.3%となり、大幅に改善できた。なお、同じ字幕の字幕画面の抽出を許容すると適合率は 87.7%となる。

CM で誤抽出が多かった原因は、画面全体で動きが多いため撮影タイミングのわずかなズレで大きな差分が生じることが挙げられる。同じ字幕の字幕画面を誤抽出した原因も、字幕文字以外に生じた差分の影響でグループ化できなかったことである。今後、字幕文字を認識することで同じ字幕かどうかの判定が可能となれば、こうした誤抽出を削減できると考える。

[参考文献]

[1] 石綿翔平, 高橋正信:“テレビ番組の字幕画面抽出”, 電子情報通信学会東京支部学生会研究発表会, 192, 2016.

[2] TBS, “BANANA FISH”, 第 1 話, 2018 年 7 月 6 日放送.