

# 動画配信におけるユーザ操作に応じたスムーズなワイプ映像切り替え手法の検討 Study on a Smooth Switching Method for Picture-in-Picture Source According to User Operation in Video Streaming

藤井 翔子<sup>†</sup> 福留 大貴<sup>†</sup> 西村 敏<sup>†</sup>  
Shoko Fujii Daiki Fukudome Satoshi Nishimura

## 1. はじめに

近年のテレビ番組では、VTR 映像等に小窓で出演者の顔映像等を合成表示するワイプ演出が広く利用されている。

しかし、放送番組でワイプ演出を実施する場合、すでに制作側によって編集された単一の動画になっているため、ユーザ側で小窓映像として表示する対象オブジェクトを切り替えたり、小窓映像の表示/非表示を選択したりすることは困難である。

そこで本稿では、新たなユーザ体験の提供に向けて、このワイプ演出を動画配信に適用し、ユーザ操作に応じたスムーズな小窓映像の切り替えを可能とする配信方式を提案する。本方式は、低容量の補完映像を利用して小窓映像の切り替え遅延を抑制すると共に、超解像技術を活用して補完映像の品質を向上する点が特徴である。

本方式を Web アプリとして実装し、性能評価を行ったので報告する。

## 2. 提案システム

### 2.1 システムの構成

提案システムの構成を図 1 に示す。

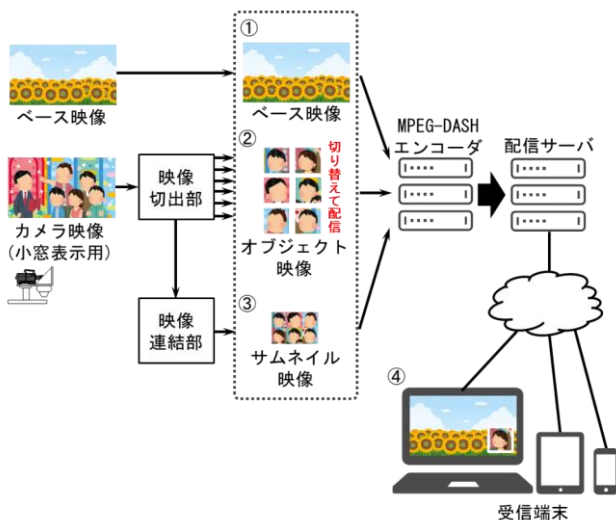


図 1 提案システムの構成

本システムでは、中継映像やVTR映像などのベースとなる映像のストリーミング視聴中に、ユーザが選択した人物の顔映像などのオブジェクトを小窓でワイプ表示する演出（図 1④）を想定する。筆者らは、これまで映像切り替えのレスポンスを向上させる手法[1][2]を提案してきたが、切り替え直後に画質が低下する課題があった。本稿では、

超解像技術の活用により、切り替え直後の画質向上を図る点が特徴である。

配信サーバには、3種類のストリームを MPEG-DASH[3]形式で生成して配置する。図 1①のベース映像は、ストリーミング視聴のメインとなる映像である。図 1②のオブジェクト映像は、小窓表示用のカメラ映像から抽出した個々の出演者の顔映像などのオブジェクトの映像であり、視聴端末上でベース映像にワイプ合成される。図 1③のサムネイル映像は、オブジェクト映像を縮小して連結させた、小窓映像の切り替えを補完する映像である。

ここで、サムネイル映像の活用方法について記述する。ストリーミング視聴において映像を切り替えることは、異なる映像をサーバに対して再度要求して、バッファリングしてから再生を行うことにあたる。そのため、オブジェクト映像を切り替えてから次の映像が再生されるまでには遅延が発生し、ユーザの体感品質が低下する恐れがある。一方で、受信端末ですべてのオブジェクト映像を同時に受信し常時デコードして表示のみを切り替えれば、切り替えの遅延は低減するが、視聴していない映像についても常時受信することになり、配信効率の低下を招く。この解決策として、オブジェクト映像を縮小しタイトル状に連結した低容量のサムネイル映像を活用する。

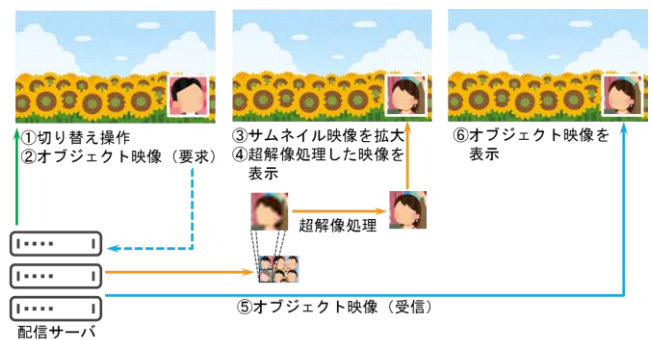


図 2 小窓映像の切り替え手順

図 2 にサムネイル映像を活用した小窓映像の切り替えの手順を示す。

ベース映像とサムネイル映像は常に配信され、オブジェクト映像はユーザ操作に応じて切り替えて配信される。ユーザが切り替え操作を行うと（図 2①）、選択したオブジェクト映像を配信サーバに要求する（図 2②）。そして、選択したオブジェクト映像が受信されるまでの間は、サムネイル映像から該当領域を切り出して表示サイズに拡大（図 2③）する。このとき、単純に拡大しただけでは品質劣化を招くため、超解像処理によって高画質化し、ベース映像の前面に表示する（図 2④）。そして、選択したオブジェクト映像を受信した後（図 2⑤）、受信したオブジェクト映像に切り替えてベース映像の前面に表示する（図 2⑥）。

<sup>†</sup> 日本放送協会 NHK(Japan Broadcasting Corporation)

このように、サムネイル映像から該当領域を切り出して表示サイズに拡大させた補完映像を利用することで配信効率と操作レスポンスの向上を図ると共に、この補完映像に対して超解像処理を実施することでストリーム切り替えまでの品質劣化を抑制することが可能となる。

## 2.2 小窓表示用ストリームの生成

オブジェクト映像の生成方法について述べる。まず、カメラ映像から小窓表示に利用する任意のオブジェクトを抽出する。本稿では、オブジェクトを顔領域に設定するため、映像の開始フレームに対して、顔検出処理によりオブジェクトを自動抽出する。そして、次のフレーム以降は、抽出したオブジェクトを、物体追跡アルゴリズムによりトラッキングし切り出すことで、オブジェクト映像を自動生成する。顔検出には顔認識ライブラリ Face Recognition[4]を、トラッキングには画像処理ライブラリ OpenCV の Tracking API[5]を使用した。

また、生成したオブジェクト映像を縮小し連結することによりサムネイル映像を生成する。

## 2.3 超解像処理

第2.1節に示すサムネイル映像の高画質化には、深層学習を用いた超解像手法 SRGAN[6]を適用する。SRGANは、敵対的生成ネットワーク (GAN: Generative Adversarial Network) を適用したモデルであり、知覚品質が高く自然であることが知られている。

ここで、SRGANは低解像度の画像と高解像度の画像の関係性を学習することにより、入力された低解像度の画像から高解像度の画像を復元する手法であり、解きたい問題に応じた適切なデータセットを用いる必要がある。そこで本稿では、復元精度の向上を図るため、配信用にエンコード済みのサムネイル映像を表示サイズに拡大した映像の各フレームを低解像度画像とし、それに対応するエンコード前のオブジェクト映像の各フレームを高解像度画像として学習を行う。

## 3. 動作検証と性能評価

提案方式を、MPEG-DASHを応用したWebアプリとして実装し、動作検証を行った。

図1のベース映像、オブジェクト映像、サムネイル映像の3種類の映像ストリームの解像度とビットレートをそれぞれ、 $1920 \times 1080/3$  Mbps,  $150 \times 150/100$  kbps,  $150 \times 50/50$  kbps (1/3に縮小した3オブジェクトの連結ストリーム)、フレームレートを30 fpsとしてH.264でエンコードを行い、セグメント長を2 sとしてMPEG-DASH化した。また、超解像処理による画質向上効果を確認するため、小窓表示用のカメラ映像を学習用と評価用に分類し、学習用映像を用いてSRGANの学習を行い、評価用映像で提案システムの性能評価を実施する。

開発した視聴プレーヤの画面例を図3に示す。ベース映像の右下を小窓表示領域とし、操作ボタンによって任意の人物名を選択することで、対象オブジェクトの切り替え操作を行う。ここで、サムネイル映像の使用有無における切り替え操作までの時間を比較する。それぞれにおいて10回切り替え操作を行った平均値を算出した。サムネイル映像を使用しない場合には、小窓映像の切り替えに約2秒を要したのに対し、提案方式では1フレーム以内の遅延に収まっ

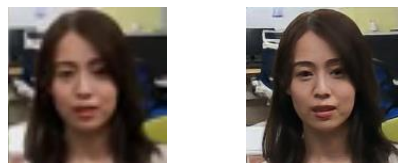


図3 作成したプレーヤの画面例

ており、ユーザ操作に応じて低遅延で切り替わることを確認した。

次に、提案システムにおける高画質化の効果を確認するため、サムネイル映像と超解像映像の画質の比較を行う。評価用映像から無作為に選択した一枚のサムネイル映像フレームに超解像処理を実施した結果を図4に示す。サムネイル映像(図4(a))と比較して、超解像映像(図4(b))の画質が向上していることが確認できる。

さらに、画質評価指標 PSNR (Peak signal-to-noise ratio) を用いて定量評価を行う。PSNRの算出にあたっては、オブジェクト映像のエンコード前のフレームを基準とし、1000フレームの平均値を算出した。サムネイル映像の PSNR が 27.36 dB であるのに対し、超解像映像の PSNR は 31.14 dB であることから、SRGAN適用によるサムネイル映像の画質向上が確認できた。



(a) サムネイル映像 (b) 超解像映像

図4 提案システムにおける高画質化の効果

## 4. おわりに

本稿では、動画配信におけるユーザ操作に応じたスムーズな小窓映像の切り替えを可能とする配信方式を提案した。

Webアプリ形式でシステムを試作し、その動作検証および性能評価により、映像品質を維持しつつレスポンス性を向上させるスムーズな小窓映像の切り替えを確認した。今後は、高画質化処理のさらなる精度向上と処理負荷の軽減を検討する。

## 参考文献

- [1] 関口頌一朗他: “MPEG-DASHを用いた多視点映像ストリーミング方式の検討”, 映像情報メディア学会年次大会講演予稿集, Vol. 2017, pp. 22D-1 (2017).
- [2] 森翔平他: “8K映像を活用した解像度を維持したズームを可能とする視聴技術の開発”, 映像情報メディア学会年次大会講演予稿集, Vol. 2019, pp. 12C-1 (2019).
- [3] 平林光浩: “次世代動画配信技術「MPEG-DASH」技術概要と標準化・関連技術動向”, 映像学誌, Vol. 67, No. 2, pp. 109-115 (2013).
- [4] Face Recognition: [https://github.com/ageitgey/face\\_recognition](https://github.com/ageitgey/face_recognition)
- [5] OpenCV Tracking API: [https://docs.opencv.org/3.4/d9/df8/group\\_\\_tracking.html](https://docs.opencv.org/3.4/d9/df8/group__tracking.html)
- [6] C. Ledig, et al.: “Photo-realistic single image super-resolution using a generative adversarial network”, Proceedings of the IEEE conference on computer vision and pattern recognition (2017).