

ライブ配信における視聴者毎の途中視聴を支援する経時的映像分割/統合システム

野崎 克[†] 橋本 浩二[†]

[†]岩手県立大学ソフトウェア情報学部

1. 概要

近年、動画配信サービスの利用の広がりとともに各種のライブ配信が実施されるようになった。その一例である e-Sports イベントのライブ配信は長時間実施される傾向にあり、途中から視聴する場合には、応援する選手のこれまでのプレイや勝敗等を短時間で把握することが難しい。一方で、長時間のライブ配信を手軽に楽しむ方法として一部の動画配信プラットフォーム [1] では配信のダイジェストを手動で生成できる。また、視聴者の興味対象となるセクションを抽出することで視聴を快適にする研究 [2] もある。しかし、視聴者の興味対象に合わせ自動で配信内容を短くまとめることはできていない。

そこで本研究では、ライブ配信の途中視聴を支援するため、視聴者毎の興味対象に合わせて映像を分割/統合することで短いダイジェストビデオを適宜自動生成し、視聴者はライブ配信のこれまでの経過を視聴開始時に容易に把握することができるシステムを提案する。本稿では、システムの概要と、機能の一部を実装したプロトタイプシステムについて報告する。

2. 提案システムの概要

以下の図 1 は提案システムの利用イメージである。

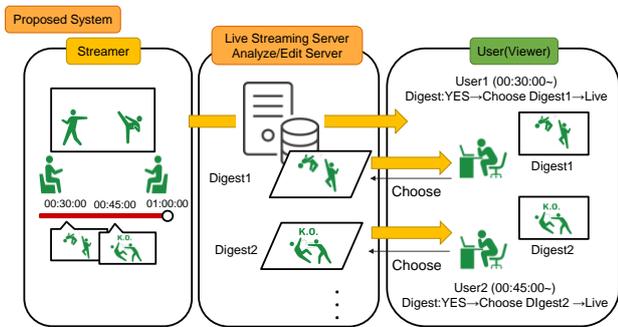


図 1 提案システムの利用イメージ

図 1 における配信者 (Streamer) は、例えば自身のゲームプレイを実況した音声とその画面をライブ配信する。配信サーバ (Live Streaming Server) はそのストリームデータを受信し、視聴者及び分析/編集サーバ (Analyze/Edit Server) へと送信する。また、システムの生成したダイジェスト映像を視聴者 (User) へ送信する。視聴者はライブ配信映像またはダイジェスト映像を選択し視聴する。

Video Segmentation/Integration System for Each User to Support for Watching from Middle in Live Streaming

Suguru Nozaki[†] and Koji Hashimoto[†]

[†]Faculty of Software and Information, Iwate Prefectural University

図 1 の User1, User2 は本研究で途中視聴者とする利用者である。User1 はライブ配信開始から 30 分、User2 は 45 分経過時点から視聴開始している。User1 は Digest1 を、User2 は Digest2 を視聴できる。ダイジェスト視聴に関して User1 は Digest1 を、User2 は Digest2 を視聴してからライブ配信を視聴している。

ここにおけるダイジェスト映像はライブ配信中に経時的に生成されるもので、つまり、経過時間によって異なるダイジェストが生成される。本研究の目的である途中視聴の補助を担う役割を持つものである。

3. システムアーキテクチャ

提案システムで必要とされる機能は大きく 4 つある。映像入出力、映像伝送、映像分析、映像分割/統合である。ハードウェア上では映像入出力は端末が、映像分析と映像分割/統合はサーバが、映像伝送は端末とサーバの双方が所持する機能と考える。

以下の図 2 に提案システムアーキテクチャを示す。

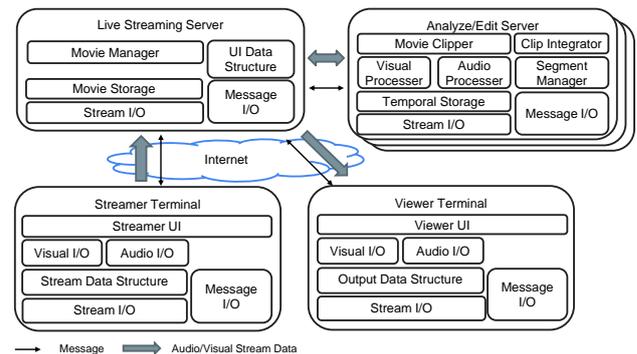


図 2 アーキテクチャ

図 2 において各 Stream I/O モジュールは映像の送受信、Message I/O はその他のメッセージを送受信する役割を持つ。また、Visual I/O 及び Audio I/O は画像/音声の入出力モジュールである。

Streamer Terminal は配信者端末であり、映像を撮影し Stream Data Structure がストリームデータ化し配信サーバへ伝送する映像データソースでもある。

Live Streaming Server は配信サーバであり配信者端末からの映像受信と視聴者端末への映像配信を行う。視聴者端末への映像伝送は視聴者端末からのメッセージを受け選択された映像を送信する。UI Data Structure は配信サーバが視聴者端末へ提供可能な映像をまとめる機能であり、これを通知するときは Stream I/O を利用する。Movie Storage はライブ配信のアーカイブ保管をする。また、後述の分析編集サーバからダイジェスト映像を受け取り保管する役割も持つ。

Analyze/Edit Server は分析/編集サーバである。受け取ったストリームデータは Temporal Storage に

バッファリングされる。画像と音声は分離され、Visual Processorが画像処理、Audio Processorが音声処理を行い、特徴抽出をして継続的に分析を実施する。分析の結果を受けてMovie Clipperがバッファリングされたデータを基に数秒から十数秒程度の短い映像を生成する。この短い映像をセグメントとする。Segment Managerはこれらセグメントを管理する。Clip Integratorが管理されたセグメントを利用しダイジェスト生成を担う。

そして、生成されたダイジェストは配信サーバへと保管され、視聴者が視聴可能な状態となる。

Viewer Terminalは視聴者端末であり、配信サーバへと視聴したい映像を通知する、あるいは配信サーバから映像を受信しOutput Data Structureが出力可能なデータに整え出力する機能を持つ。

4. 経時的映像分割/統合機能

提案システム中でセグメント及びダイジェストの生成に係る機能を説明する。ここで、検出可能な視聴者の興味対象を便宜上トリガーと呼ぶことにする。

以下の図3は提案システムにおける経時的映像分割/統合機能の概要図である。

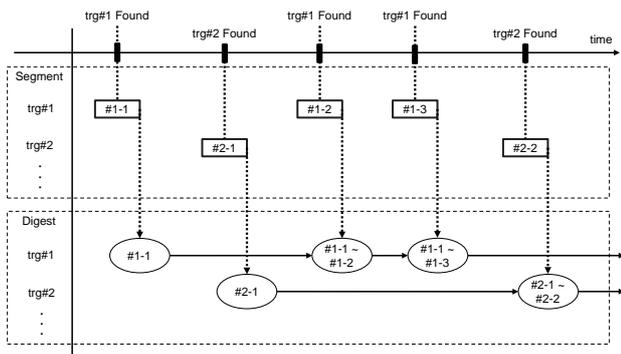


図3 映像分割/統合機能概要

図3ではトリガーtrgの検出によって生成される映像を時間軸と共に端的に表している。矩形Segment中の#n-mはtrg#nで生成されたm番目のセグメントという意味である。矩形Digest中の#n-m~#n-M(M:最終セグメント番号)はセグメント#n-m~#n-Mを収集しまとめたダイジェスト映像となっている。下記は大まかな流れである。

画像/音声の分析でtrg#nが検出されたとき、検出位置周辺の映像をセグメント#n-mとして抽出し保存する。セグメントの生成時点でそれと同じトリガーによるセグメントを収集し、ダイジェスト#n-m~n-Mが生成される。トリガーの種類ごとにダイジェストも種別で保管される。新たにtrg#nのセグメントが表れるまで、ダイジェスト#n-m~n-Mが最新のダイジェストとして視聴者に提供される。

この機能概要に基づいたプロトタイプの実装を次項から述べていく。

5. プロトタイプ実装

今回のプロトタイプは分析/編集サーバ及び視聴者端末の機能を持つ。実装において分析モジュールで利用する視聴者の興味対象は、個人差があり一概にまとめることは難しい。よって、担当人員が動画を

視聴し分析するUIとして実装した。さらに、このUIでの分析結果に基づいてセグメントを生成しダイジェスト化するモジュールを実装した。これにより視聴者の興味対象となり得る要素を疑似的ながら経時的にまとめることができる。以下は実装したプロトタイプタイプの概略図である。

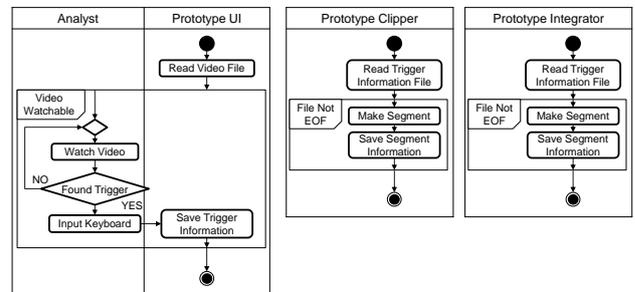


図4 プロトタイプの概略図

図4のプロトタイプとアーキテクチャの対応を以下に示す。Analystは分析担当人員を示し、図2におけるAudio Processor及びVisual Processorの代替である。Prototype UIはViewer TerminalのVisual I/O及びAudio I/Oの機能を持つ。Prototype Clipperはセグメント生成を、Prototype Integratorはダイジェスト生成を行うモジュールで、それぞれMovie ClipperとClip Integratorの機能である。

分析は担当人員の主観により行い、担当人員のボタン入力をトリガーとする。セグメント生成及びダイジェスト化は自動的に行われる。セグメント長は5~15秒である。ダイジェスト化は同系統セグメントのみ収集する方針と、経過時間に沿い周期的にセグメントを収集する方針で行う。

このプロトタイプを使用し、録画映像を用いてセグメント生成及びダイジェスト生成が行えることを確認した。引き続き、これらダイジェスト映像を用いた評価実験を計画している。途中視聴となるように仮想の本配信として録画映像の後半部分を用意する。そして、被験者に生成されたダイジェストと仮想本配信を視聴させアンケートに回答してもらう形で実施予定である。

6. まとめ

本稿では、途中視聴補助を目的とした経時的映像分割/統合システムを提案した。ストリームデータを利用した視聴者の興味対象による映像分割/統合機能のプロトタイプを実装した。今後は、視聴者の興味対象によるダイジェストのライブ配信コンテンツへの理解貢献度を評価するため、評価実験を実施する予定である。

7. 参考文献

[1]Twitch <https://www.twitch.tv/> (参照 2022/01/04).
 [2]C. Ailie Fraser et al. 2020. Temporal Segmentation of Creative Live Streams. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), April 25–30, 2020, Honolulu, HI, USA, Pages 1–12. <https://doi.org/10.1145/3313831.3376437>