Vol.2021-HCI-191 No.18
2021/1/29

多視点カメラ映像の選好を収集するための視聴行為を阻害しない映像提示方法の検討

遠藤 史央里[1]　竹川 佳成[1]　松村 耕平[2]　平田 圭二[1]　五十嵐 健夫[3]

概要：本研究では，複数のカメラを用いて撮影された多視点映像から適切な映像を選択するスイッチングを自動化するためのデータセットの構築を効率的に行うためのインタフェースを提案する．従来，複数のカメラ映像の中から適切な映像を選択するタスク（スイッチング）は，手動で行われてきた．我々はこのスイッチングについて，機械学習による自動化を目指している．この自動化のためには多視点映像の中から選好される映像がアノテーションされたデータセットを構築する必要がある．本稿では，このようなデータセットの構築に向け，視聴行為を阻害せずにアノテーション数を増加させる映像提示方法を提案する．具体的には多視点映像を閲覧するためのビューアにおいて一定時間でカメラ映像が自動的に切り替わるインターフェースを開発した．比較調査の結果，カメラ映像が自動で切り替わる提案法は，その機能を持たない従来法に比べて映像の内容理解やアノテーション負荷の低さに差はなかった一方，アノテーション回数が有意に増加したことが明らかになった．

1. はじめに

近年，YouTube やニコニコ動画を始めとする動画共有サービスや Netflix や Hulu などの動画配信サービス上の映像コンテンツは爆発的に増加している．専門知識を持たずともスマートフォン一つあれば映像を配信できるという手軽さから，プロフェッショナルからアマチュアまで映像クリエータの裾野は拡がっている．同時に，民生用のカメラや映像ミキサが開発され，アマチュアクリエータが複数のカメラを用いて多視点カメラ映像を撮影し，編集し，配信するケースも増えている．配信映像コンテンツは音楽ライブ，演劇，講義，e スポーツ，学会発表など多岐にわたる．

多視点カメラ映像を編集することで映像コンテンツの品質を高められる一方，データ量が膨大になり手動による編集作業には限界がある．このため，映像に対してアノテーションを付与し，これらのデータセットを訓練データとして機械学習アルゴリズムに適用しモデルを構築することで，映像検索支援，映像要約支援，映像編集支援など機械学習応用システムを開発できる．機械学習応用システムの動作は，アルゴリズムだけでなく訓練データの品質に依存するため，訓練データを適切に準備することが重要である．

本研究では，多視点カメラ映像におけるアノテーションデータセット構築の第一段階として，複数のカメラ映像の内どのカメラ映像を視聴したいかという選好データを収集するための映像提示方法の構築をめざす．具体的には，シンプルなカメラスイッチングインタフェースを構築し，スイッチング結果が反映される画面（以降，メイン映像とする）へのカメラ映像の表示方法として，カメラ映像の自動切り替え（被験者が自らカメラ映像を選択しない場合，メイン映像に表示されているカメラ映像とは別のカメラ映像に自動的に切り替わる機能）や blank 映像[*1]の挿入を導入した．提案する映像提示方法の有用性を検証するための被験者実験を実施した．その結果，カメラ映像の自動切り替えおよび blank 映像の挿入は，比較手法と比較して映像の内容理解やアノテーション負荷の低さに差はなかった一方，アノテーション回数が有意に増加したことが明らかになった．

多視点カメラ映像の選好に関するデータセットを構築できれば，CM 挿入タイミングの自動化，まとめ動画の自動生成，スイッチング業務の自動化などさまざまなアプリケーションを開発できる．また，視聴者が視聴したい映像は，シーン全体の文脈理解や，あるシーンの意味理解を必要とし，既存の画像処理技術を使ったアノテーションは難しく，意味理解が必要な要素技術の発展に貢献できる．

2. 関連研究

斎藤らは，動画中の特定のシーンと関連付けられたコメ

[1] 公立はこだて未来大学
[2] 立命館大学
[3] 東京大学

[*1] 何も映っていない黒の映像

ント数に着目し，シーンの特徴を推定することで，動画の視聴を妨げない広告動画の適切な挿入タイミングを推定している [17]．この研究は，投稿されたコメントを時系列で分析し単位時間ごとのコメント数に着目している．本研究では，複数のカメラ映像から選択された単位時間ごとのスイッチング回数に着目している点が異なる．

アノテーションシステムを用いた研究がいくつかある．Kovacs はユーザが十分理解できていない問いを提示するシステム [5] を，Bargeron らは個人的なメモ作成やメモの共有のための共同ビデオ注釈システム [1] を，Costa らは同じビデオコンテンツに対していくつかの視点を提供し複数のビューで表示できる設計のシステム [3] を構築している．Weher ら [16] は，ペンベースのビデオアノテーションツールを開発している．これにより，ユーザーは記録中にメモやキーワードをビデオテープに関連付けることができる．また，特定のカテゴリのビデオ用のアノテーションツールが提案されている．例えば，Utasi ら [10] は人の検出，Cabra ら [2] はコンテンポラリーダンス用，Miller ら [8] は動くターゲットとハウツービデオ用のツールを提案している．これらの研究は全て，一つのコンテンツとして完成された映像に対するアノテーションを対象としている．本研究は，複数のカメラで撮影された映像を対象としている点で異なる．

また，多視点カメラの自動スイッチングに関する研究もいくつか存在する．Wang らはサッカー用の自動カメラスイッチングシステムを提案している．ボールやプレイヤの位置 [11], [12]，軌跡 [14], [15]，視聴者の興味 [13] に基づいて各カメラのスコアを計算する研究である．Leake ら [6] は対話シーンのビデオを編集するためのシステムを提案している．システムはユーザーが指定した映像編集イディオムのセットに基づき，複数のカメラから最も適切なクリップを自動的に選択する．土田らはダンスを対象に，多視点カメラ映像から特徴量を抽出し，特徴量に対応付けられたスライダーバーを操作することで，ダンス映像制作者が好みの映像を作るシステムを提案している [9]．松井らは，ピアノレッスンを円滑に進めるために，多視点カメラ映像の各映像フレームにタグ付けされたデータに対して特徴量を抽出し，機械学習を適用することで，カメラ映像の自動切り替えを実現している [7]．長谷川らは，ピアノ演奏の指使いに関する悪癖を発見することを目的とし，多視点映像を効率的に閲覧できるインタフェースを提案している [18]．これらの研究はいずれも対象を限定しその対象に特化した GUI を提案しており，今回の目的に即したインタフェースではない．

3. 設計と実装

本研究の目的は，視聴を阻害せず自然かつ効率的に多視点カメラ映像の選好を収集するための映像提示方法の提案である．想定している利用シーンとして，自宅のデスクといった集中できる環境で十分な時間や余裕がある状況で映像を視聴する以外に，電車やバスでの移動中など，映像視聴は可能だが，メモをするなど複雑な操作ができないような状況も想定している．また，短い動画であるほどエンゲージメント率が高くなる[*2]という動画マーケティングの結果が得られており，アノテーションのために何度も映像を見直すなどで映像を視聴する時間が長くなってしまうことは避けるべきである．さらに，スマートフォンを利用した動画視聴率は 9 割を超えている[*3]という統計データをもとに，スマートフォンを利用した動画視聴を想定する．これらを前提とし，以下の要件をバランスよく満たす必要がある．

要件

- アノテーション回数の多さ：ユーザの明確な映像の選択の意思をアノテーションとして多く取得する必要がある．
- 映像の内容への理解度の高さ：上述したようにユーザは自分が見たい映像コンテンツを視聴するということを前提としているため，ユーザがアノテーションに集中しすぎる余り，どういう映像を見ていたか内容を理解できなくなってしまうことは避けるべきである．
- アノテーション作業への負荷の低さ：映像の視聴がユーザにとって主目的であるため，アノテーション作業に負担がかからないようにする必要がある．
- 必要最低限の拘束時間：電車の移動中に動画の視聴を完結させたいといったように，アノテーション作業による動画視聴時間の延長は避けるべきである．

3.1 インタフェース

提案インタフェースのスクリーンスナップショットを**図 1** に示す．提案インタフェースは，一般的な動画再生アプリケーションと同様の画面構成であるが，多視点カメラ映像であるため，サムネイル映像が表示されているという点が大きく異なる．スマートフォンでの利用を想定しているため，表示サイズは一般的なスマートフォンのサイズに近い縦 75mm，横 150mm の範囲に収められている．ページの左上隅の位置に，各カメラ映像のパネルが縦に並べて表示される．各カメラ映像の右隣には，カメラ映像の内のいずれか一つがメイン映像として大きく表示される．メイン映像の下部に再生ボタンと経過時間を示すシークバーが表示されており，再生ボタンをクリックすることで発表映像の再生が開始される．ユーザがカメラ映像のパネルのうちいずれかをクリックで選択すると，メイン映像が選択したカメラ映像へと切り替わる．

提案インタフェースのメイン映像のように比較的大きめ

[*2] https://wistia.com/learn/marketing/optimal-video-length
[*3] https://webtan.impress.co.jp/n/2019/01/07/31478

図 1 インターフェースのスクリーンスナップショット

の表示を用意せず，各カメラ映像すべてを同じ大きさで表示させた場合，スマートフォンの表示サイズの限界から各映像の細部が見えにくくなってしまう．また，カメラ映像の選好データの収集を目的としているが，このような状態で，視聴者に視聴したい映像を選択してもらう場合，視聴者に選択するインセンティブを提供できない．したがって，各カメラ映像の表示サイズは小さく，選択された映像が映るメイン映像の表示サイズは大きくなるようにした．各カメラ映像や再生ボタンは画面左側に固めて配置しているため，片手での操作も容易となっている．

3.2 映像提示方法

本研究で提案する映像提示方法として，映像の自動切り替え，および，blank 映像の挿入を導入する．

映像の自動切り替えとは，スマートフォン上のメイン映像に表示されるカメラ映像が，自動的に切り替わる機能である．ある時点で見たい映像ではないカメラ映像がメイン映像に表示されているとき，ユーザは自然に視聴したいカメラ映像を選択すると期待される．なお，本研究では，先行研究 [4] におけるプロフェッショナルスイッチャのスイッチング分析結果をもとに自動切り替え時間を 5 秒と設定した．

blank 映像の挿入とは，メイン映像に blank 映像を挿入する機能である．メイン映像に blank 映像が表示された場合，ユーザは反射的に視聴したいカメラ映像を選択すると期待される．また，仮に blank 映像が表示されている最中にカメラが選択されなかった場合，例えば，ユーザにとって映像を見なくても音声だけで理解できると解釈でき，このようなシーンがわかれば，広告の挿入タイミングなどに活用することもできる．

その他の映像提示手法として，一定時間経過すると動画の再生が停止し，視聴したいカメラ映像を選択すると動画が再生するような手法も検討した．この手法であれば，確実にカメラ映像を選択してもらえるが，アノテーション作業を含む動画視聴時間が長くなってしまうため，導入しなかった．

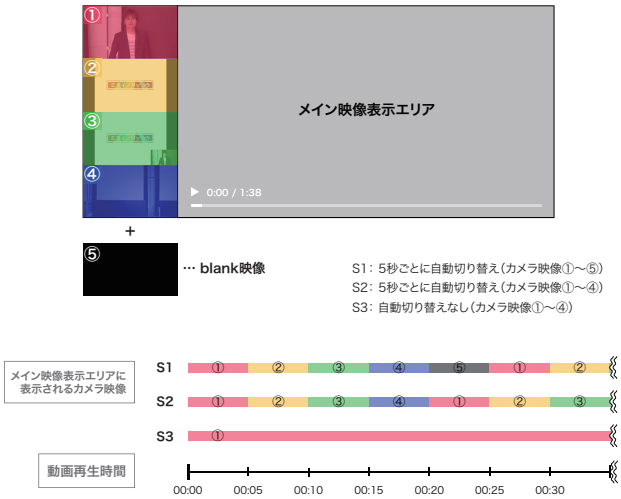

図 2 映像の提示方法

4. 実験

3 章で説明した映像提示方法および提案インタフェースの有用性を検証するために，複数台のカメラで撮影された発表映像を視聴する実験を実施した．

映像の提示方法は，**図 2** に示すとおり，自動切り替え（4 つのカメラ映像および blank 映像），自動切り替え（4 つのカメラ映像），自動切り替えなしの 3 種類とした．自動切り替えなしが比較手法である．また，自動切り替え（4 つのカメラ映像および blank 映像）と自動切り替え（4 つのカメラ映像）とを比較することで blank 映像の挿入の効果を検証できる．以降，それぞれの方法を S1，S2，S3 と呼ぶ．S1 では，5 秒ごとに 4 つのカメラ映像と blank 映像の計 5 つの映像を切り替えてメイン映像に表示する．S2 では，5 秒ごとに 4 つのカメラ映像を切り替えてメイン映像に表示する．S3 では，カメラ映像を自動的に切り替えずに提示する．

各提示方法において，映像を視聴しながら見たいカメラ映像を選択してもらうというアノテーションタスクを被験者に取り組んでもらう．当該タスク中に生じるアノテーションを記録し分析することで，提案手法がアノテーションデータの回数の増加に貢献できるか検証する．

また，本実験では，カメラ映像を選択するというアノテーション作業に対する負荷について被験者へのアンケートによる主観評価（5 段階評価：1:負荷が高い〜5:負荷が低い）で評価した．また，内容への理解度についてはアンケートによる主観評価（5 段階評価：1:理解度が低い〜5:理解度が高い）のほか，簡単な理解度テスト（選択式全 2 問）で評価した．提案インタフェースは視聴行為を阻害するかどうかを，理解度や作業負荷という評価指標で評価する．

4.1 発表映像

被験者が視聴する発表映像は 3 種類あり，それぞれ約 1

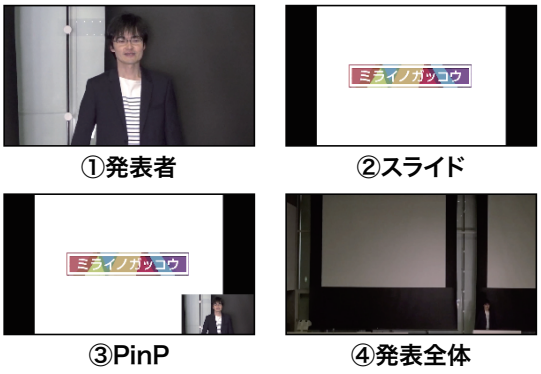

図 3 各カメラ映像

表 1 手法ごとのサンプルサイズ

提示方法	サンプルサイズ
S1：自動切り替え（4 カメラ映像＋ blank 映像）	63
S2：自動切り替え（4 カメラ映像）	55
S3：自動切り替えなし	59

分 30 秒程度の長さであった．情報処理学会が主催している IPSJ-ONE やインタラクションなど多くの学会発表の中継では，4 つのカメラ映像を切り替えて配信している．したがって，本実験における各発表映像のいずれも 4 つのカメラで撮影したものを使用した．各カメラ映像を**図 3** に示す．カメラ映像はそれぞれ，発表者・スライド・PinP（スライドの右下に発表者）・発表全体であった．被験者は，インタフェースに表示された各カメラ映像のパネルをクリックすることでカメラ映像を選択した．

4.2 被験者

被験者は，映像・撮影に関して専門的な知識をもたないカメラスイッチングの初心者 177 人であった．本実験は被検者間実験を取り入れ，1 人の被験者に S1〜S3 のいずれか 1 つの提示方法を割り当てた．各手法ごとのサンプルサイズを**表 1** に示す．本来であれば，カウンターバランスをとるために，手法ごとのサンプルサイズは等しくすべきであるが，欠損データなどがあり，サンプルサイズは手法ごとに異なった．しかし，分析をする上で十分なサンプルサイズを確保できていると考えたため，サンプルサイズの違いは結果に影響しないと考えている．

また，被験者は自室のデスク上といったように実験に集中できる環境で実験に取り組んでもらった．また，被験者は提案インタフェースや視聴してもらう発表映像に対して事前知識を持ち合わせていない．

4.3 実験の手続き

最初に，被験者に 30 秒程度の練習用の発表映像を視聴してもらい，インタフェースの操作方法を理解してもらった．実験者は被験者に視聴方法やインタフェースの操作方法に問題がないか確認した．被験者には，「4 つのサムネイル映像からその都度自分が見たい映像を選択してください」と指示した．その後，被験者には 3 種類の映像からランダムで 1 つの映像を視聴してもらい，割り当てられた映像を視聴した後，理解度確認テストやアンケート（発表内容への理解度や，アノテーション作業の負荷）に回答してもらった．

4.4 実験結果

各手法の平均アノテーション回数を**図 4** に示す．回数が多かったものから順に，S1，S2，S3 であった．分散分析を適用したところ，S1 および S3 間，S2 および S3 間で有意水準 5%において有意差が観測された（$F(2, 174) = 8.54$, $p < .05$）．

選択されたカメラ映像の分布を**図 5** に示す．S1，S2 では選択された割合が高かったものから順に，PinP，スライド，発表全体，発表者であり，S3 では選択された割合が高かったものから順に，スライド，PinP，発表全体，発表者であった．また，選択された映像の品質について調査するために，文献 [4] におけるプロフェッショナルのスイッチング条件をもとに作成した正解データの分布を図 5 の右端の円グラフに示す．選択された割合が高かったものから順に，PinP，スライド，発表全体，発表者という結果になった．

理解度テストの平均点を**図 6** に示す．理解度テストでは平均点が高かったものから順に，S1，S2，S3 であった．有意水準 5 %においてクラスカル・ウォリス検定により検定したところ，有意差は観測されなかった．また，発表内容への理解度を主観的に回答してもらった平均スコアを**図 7** に示す．5 段階評価の平均値は S1 と S3 が同じスコアとなり，次いで S2 となった．有意水準 5 %においてクラスカル・ウォリス検定により検定したところ，有意差は観測されなかった．

アノテーション作業の負荷について主観的に回答してもらった平均スコアを**図 8** に示す，負荷が低かったものから順に，S1, S2, S3 であった．有意水準 5 %においてクラスカル・ウォリス検定により検定したところ，有意差は観測されなかった．

4.5 考察

アノテーションの回数

アノテーション回数について，図 4 の平均アノテーション回数を見ると，自動切り替えのある S1，S2 は自動切り替えのない S3 よりも多くアノテーションされていることがわかる．自動切り替えは，アノテーションを活性化する効果を高くすることが明らかになったといえる．また，自動切り替えでは blank 映像の挿入がある S1 は blank 映像の挿入がない S2 よりも多くアノテーションされていることがわかる．S1 および S2 間に関して，検定の結果より有意

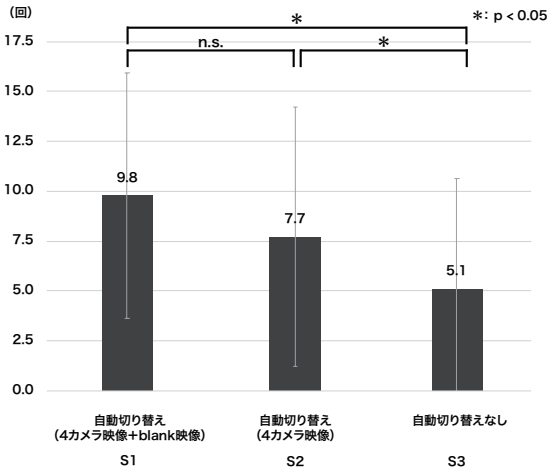

図 4　平均アノテーション回数

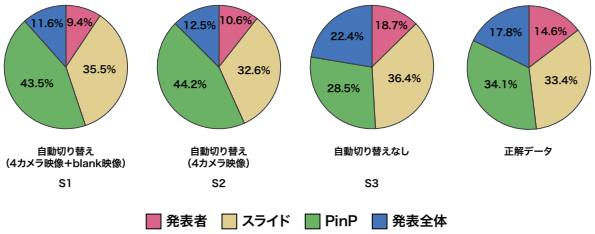

図 5　選択されたカメラ映像の分布

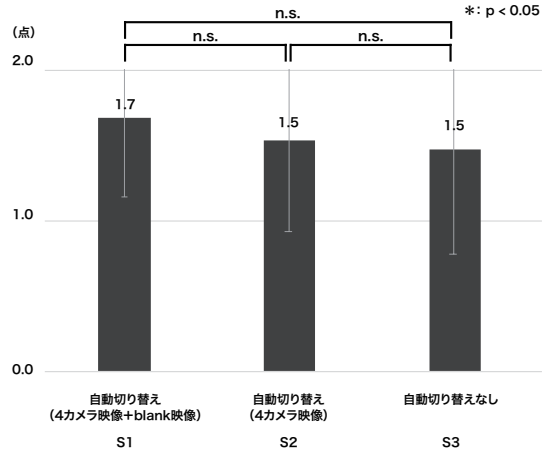

図 6　理解度テストの平均点（2 点満点）

差は見られなかった．このことから，blank 映像の挿入そのものがアノテーションの活性化に対して高い効果があるとはいえないが，効果の高い自動切り替えに加えて blank 映像の挿入を加えると，自動切り替え単体で提示するよりはアノテーション回数を増やすことが可能であると考えられる．

自動切り替えが発生した際に自動切り替え発生前のカメラ映像に戻した割合（戻し発生率と呼ぶ）の平均値は 19.3%となり，戻し発生率は低くなった．自動切り替え時間が短すぎると，元のカメラに戻す，あるいは，何も選択されずに放置されてしまう恐れがある．S1 あるいは S2 ではスイッチング回数が多くなったこと，また，平均戻し発生率も低

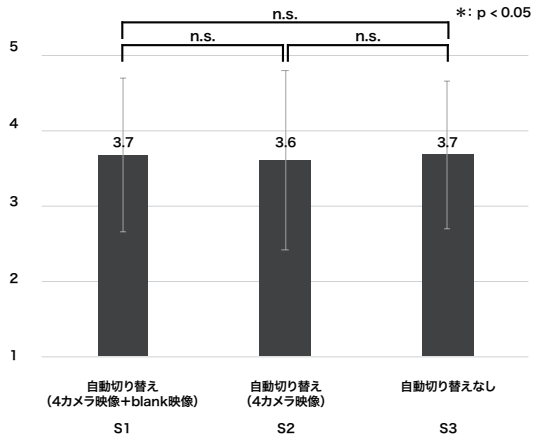

図 7　発表内容への理解度
（1:理解度が低い〜5:理解度が高い）

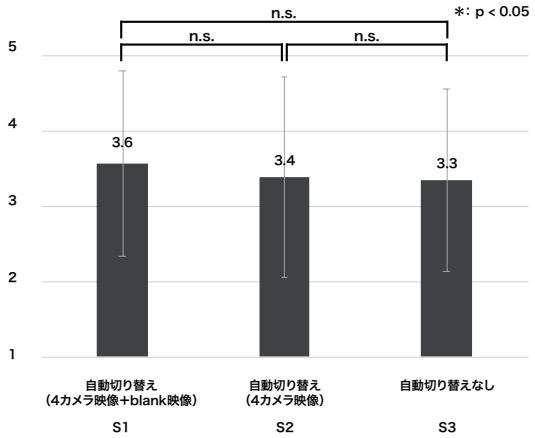

図 8　アノテーション作業への負荷の低さ
（1:負荷が高い〜5:負荷が低い）

かったことから，提案手法で設定した 5 秒間は，現在見ている映像の内容を理解し，別のカメラ映像を選択するかどうか判断する上で妥当な時間設定だったといえる．

blank 映像が挿入されているタイミングのみに着目し，その区間でアノテーションが発生しているかを調査した．blank 映像への反応率（blank 発生時のアノテーション発生数/blank 発生数）の平均値は 71.4%となった．blank 映像がメイン映像として表示されたとき，「カメラ映像を選択したくなる」というコメントをもらい，blank 映像の提示は概ね意図した通りの反応が観測された．一方，blank 映像時に切り替えなかった場合も散見され，「音声が聞こえているので黒い画面でも理解可能であったから」「待っていればそのうち映像がまた切り替わるから」というコメントが少なからずあった．このように映像中のある区間において映像そのものが不要なため選択されなかった．

アノテーションの分布

図 5 に示す選択されたカメラ映像の分布について，S1 から S3 および正解データいずれも，発表全体や PinP が選択された割合は多く，発表者や発表全体が選択された割合は少

なくなった．この結果より，各手法は正解データのカメラ映像の分布と類似しており，アノテーションの品質という点で一定の保証されたと考えられる．S1 や S2 で PinP の割合が高くなった理由については現段階では調査できていないため，今後の課題とする．

理解度

発表内容への理解度ついて，図 6 より，理解度テストにおいて 3 つの手法間で平均点に差は見られなかった．また，図 7 より，発表内容への理解度に関するアンケート結果においても 3 つの手法間で平均値に差は見られなかった．いずれの手法も理解度テストの平均点は高く，理解度に関するアンケート結果の平均スコアも 3.6 あるいは 3.7 となり，アノテーション作業をしながらでも発表内容は理解できていたといえる．S1 や S2 などの自動切り替えがある手法について，被験者自身が特に見たいと考えていないカメラ映像へと勝手に切り替えられることにより，映像の内容への理解度に悪影響が及ぶ可能性が考えられた．さらに，S1 では blank 映像が挿入され，視覚情報が失われるため，被験者による映像の内容への理解が阻害されることを危惧していた．しかし，理解度テストやアンケートによる主観評価の結果より，どの映像提示手法に置いても理解度に差は現れないことが明らかになったため，理解度を保ちつつアノテーション数を多く取得するために自動切り替えや blank の挿入は有用であると考えられる．

負荷

アノテーション作業の負荷について，図 8 より，3 つの手法間で平均値に差は見られなかった．S3 のように被験者自身のタイミングでカメラ映像を選択する手法と比較して，S1 や S2 の手法では見ている映像が被験者の望まないタイミングで切り替わる可能性がある．それに伴い，カメラ映像を逐一戻す労力が必要とされることで，被験者はアノテーションへの負荷を高く感じることが予想された．しかし，アンケートによる主観評価の結果より，どの映像提示手法に置いてもアノテーションへの負荷に差はないことが明らかになり，自動切り替えや blank の挿入をしても被験者への負荷にはならないことが示唆された．

4.6 本実験の限界

本論文の評価実験では，被験者が視聴する映像を発表映像に限定して実施したため，他のコンテンツに対しては同様の結果が得られない可能性がある．対象とする映像の幅を広げて新たに実験を実施し，検証する必要がある．また，映像全体を通してアノテーションデータの量を増やすことに関して提案手法の有意性は確認でき，アノテーションデータの品質に関して正解データと大きな差はないことが確認できた．しかし，ある時点や区間におけるアノテーションデータの量および質に関しては議論できていない．実験データの詳細な分析は今後の課題である．

本実験は，3 章の冒頭で説明した利用シナリオ（電車やバスでの移動中）での実験は実施しておらず，自室のデスク上というノイズが少なく落ち着いた状況で実験を実施した．電車やバスなどでの移動中の場合，片手で操作をしないといけなかったり，振動などで画面が揺れたり，周囲に気を配らなければいけなかったりなど様々な外乱が生じると考えられる．このような状況下における提案手法の有用性に関する評価は今後の課題である．

5. まとめ

本稿では，機械学習を通じたスイッチングの自動化を目指し，多視点カメラ映像をブラウザで視聴できるインタフェースを構築し，効率的なデータ構築に向けたアノテーション収集の仕組みとして視聴行為を阻害しない映像提示方法を検討した. インターフェースを用いて実施した発表映像の視聴実験を通じ，自動切り替えや blank 映像挿入によってアノテーションを多くすることが可能であると明らかになった．自動切り替えや blank 映像の挿入がある映像提示方法は，通常の映像提示方法と比較してもユーザによる映像の内容への理解度に影響はなく，またカメラ映像を切り替えながらの視聴行為に対する負荷も低かった．選択されるカメラ映像の特徴やタイミング等については，さらに議論を重ね検討していく必要がある．

今後は，映像の提示条件や評価方法を検討しつつ，映像の視聴によるアノテーションデータを継続して収集する．通常，学会発表は 1 発表につき 10 分以上の時間を有することが多いため，10 分以上の長さがある映像を用いた実験も実施する必要がある．さらに，効率的なデータセットの構築をめざし，一定のデータが蓄積された状況で，データ数が少ないシーンに対して動的に blank 映像を挿入するアルゴリズムの構築などがあげられる．

謝辞　本研究は JST CREST JPMJCR17A1 の支援を受けたものである．

参考文献

[1] Bargeron, D., Gupta, A., Grudin, J. and Sanocki, E.: Annotations for Streaming Video on the Web, *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '99, New York, NY, USA, Association for Computing Machinery, p. 278?279 (online), DOI: 10.1145/632716.632887 (1999).

[2] Cabral, D., Carvalho, U., Silva, J. a., Valente, J. a., Fernandes, C. and Correia, N.: Multimodal Video Annotation for Contemporary Dance Creation, CHI ' 11 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, Association for Computing Machinery, p. 2293–2298 (online), DOI: 10.1145/1979742.1979930 (2011).

[3] Costa, M., Correia, N. and Guimarães, N.: Annotations as Multiple Perspectives of Video Content, *Proceedings of the Tenth ACM International Conference on Multimedia*, MULTIMEDIA '02, New York, NY, USA, Association for Computing Machinery, p. 283?286 (online), DOI: 10.1145/641007.641065 (2002).

[4] Endo, S., Takegawa, Y., Funaki, A., Matsumura, K., Hirata, K. and Igarashi, T.: Construction of a Switching Support System for Live Broadcast of Oral Presentation, *Journal of Information Processing*, p. to appear (2021).

[5] Kovacs, G.: QuizCram: A Question-Driven Video Studying Interface, *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, New York, NY, USA, Association for Computing Machinery, p. 133?138 (online), DOI: 10.1145/2702613.2726966 (2015).

[6] Leake, M., Davis, A., Truong, A. and Agrawala, M.: Computational Video Editing for Dialogue-Driven Scenes, *ACM Trans. Graph.*, Vol. 36, No. 4 (online), DOI: 10.1145/3072959.3073653 (2017).

[7] Matsui, R., Takegawa, Y. and Hirata, K.: Remote Piano Lesson System Considering Camera Switching, *Proceedings of International Computer Music Conference*, ICMA, pp. 1–7 (2019).

[8] Miller, G., Fels, S., Al Hajri, A., Ilich, M., Foley-Fisher, Z., Fernandez, M. and Jang, D.: MediaDiver: Viewing and annotating multi-view video, *CHI EA 2011 - 29th Annual CHI Conference on Human Factors in Computing Systems, Conference Proceedings and Extended Abstracts*, pp. 1141–1146 (online), DOI: 10.1145/1979742.1979711 (2011).

[9] Tsuchida, S., Fukayama, S. and Goto, M.: Automatic System for Editing Dance Videos Recorded Using Multiple Cameras, *Advances in Computer Entertainment Technology* (Cheok, A. D., Inami, M. and Romão, T., eds.), Cham, Springer International Publishing, pp. 671–688 (2018).

[10] Utasi, A. and Benedek, C.: A Multi-View Annotation Tool for People Detection Evaluation, *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/2304496.2304499 (2012).

[11] Wang, X., Hirayama, T. and Mase, K.: Viewpoint Sequence Recommendation Based on Contextual Information for Multiview Video, *IEEE MultiMedia*, Vol. 22, No. 4, pp. 40–50 (online), DOI: 10.1109/MMUL.2015.75 (2015).

[12] Wang, X., Muramatu, Y., Hirayama, T. and Mase, K.: Context-Dependent Viewpoint Sequence Recommendation System for Multi-view Video, *2014 IEEE International Symposium on Multimedia*, pp. 195–202 (online), DOI: 10.1109/ISM.2014.44 (2014).

[13] Wang, X., Enokibori, Y., Hirayama, T., Hara, K. and Mase, K.: User Group Based Viewpoint Recommendation Using User Attributes for Multiview Videos, *Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes*, New York, NY, USA, Association for Computing Machinery, p. 3–9 (online), DOI: 10.1145/3132515.3132523 (2017).

[14] Wang, X., Hara, K., Enokibori, Y., Hirayama, T. and Mase, K.: Personal Multi-View Viewpoint Recommendation Based on Trajectory Distribution of the Viewing Target, *Proceedings of the 24th ACM International Conference on Multimedia*, New York, NY, USA, Association for Computing Machinery, p. 471–475 (online), DOI: 10.1145/2964284.2967265 (2016).

[15] WANG, X., HARA, K., ENOKIBORI, Y., HIRAYAMA, T. and MASE, K.: Personal Viewpoint Navigation Based on Object Trajectory Distribution for Multi-View Videos, *IEICE Transactions on Information and Systems*, Vol. E101.D, No. 1, pp. 193–204 (online), DOI: 10.1587/transinf.2017EDP7122 (2018).

[16] Weher, K. and Poon, A.: Marquee: A Tool for Real-Time Video Logging, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, Association for Computing Machinery, p. 58–64 (online), DOI: 10.1145/191666.191697 (1994).

[17] 齊藤義仰，村山優子：視聴者コメントを用いた広告動画挿入タイミング決定アルゴリズムの提案と評価，情報処理学会論文誌，Vol. 52, No. 2, pp. 520–528（オンライン），入手先 ⟨https://ci.nii.ac.jp/naid/110008507893/⟩ (2011).

[18] 松井遼太，長谷川麻美，竹川佳成，平田圭二，柳沢　豊：ピアノ教師向け悪癖発見支援システムの設計と実装および評価，情報処理学会論文誌，Vol. 61, No. 4, pp. 789–797 (2020).