Vol.2019-GN-107 No.11
2019/3/19

360 度インターネット生放送における POV 追従検出手法の提案

高田真也 [1]　西岡大 [1]　齊藤義仰 [1]

Proposal of a Method for Detection of POV Tracking in 360-degree Internet Live Broadcasting

MASAYA TAKADA[1]　DAI NISHIOKA[1]　YOSHIA SAITO[1]

1. はじめに

インターネット生放送サービスは YouTube をはじめとして様々な Web サイトで提供されており，多くのユーザによって利用されている．インターネット生放送において，放送者は視聴者とコメントを用いたリアルタイムコミュニケーションを楽しむことができる．現在，YouTube では全方位カメラに対応した 360 度インターネット生放送サービスの提供も開始されており，一般ユーザも全方位カメラを用いた放送を利用することができるようになった．360 度インターネット生放送とは，全方位カメラによって撮影される全天球映像を視聴できる放送形式であり，視聴者は自身の興味に合わせて POV を 360 度全方向に変更することができる．POV とは Point of view の略称で，視聴者が視聴している方向を示す．

しかし，360 度インターネット生放送では，放送者は視聴者の POV を確認する方法がないため，視聴者の興味関心が把握できないという問題がある．従来のインターネット生放送では，単一のレンズを持った Web カメラが用いられており，単一のレンズが限定的な撮影範囲を示す．そのため，放送者は感覚的に視聴者の視聴している映像を把握することができた．それに対し，360 度インターネット生放送では，魚眼レンズや複数のレンズを搭載した全方位カメラを用いるため，視聴者の視聴している範囲がどこなのか放送者は把握することができなくなった．これまでにもコミュニケーションにおいてコミュニケーション相手の注視情報や視線情報が担う役割については多く論じられている[1][2]．その中で，コミュニケーション相手の注視情報は相手の興味や関心，話題の中心を示すということが明らかにされている．つまり，360 度インターネット生放送において，視聴者の POV は視聴者が全方位映像の中で何を見ているのか示す情報であると同時に視聴者の興味や関心を示す情報でもある．

我々はこれまでこの問題に対して球体状のヒートマップを用いて視聴者の POV を可視化する手法を検討してきた[3]．しかし，これまで行ってきた検討では，全方位カメラを固定したユースケースのみに着目しており，可搬式の全方位カメラを用いた 360 度インターネット生放送については検討できていなかった．360 度インターネット生放送は従来の放送形態と比較して撮影範囲が広く，より多くの情報を視聴者に提供することができる．そのため，観光地の紹介など，空間全体の情報を提供する目的で利用されることが想定できる．具体的なユースケースとして，放送者が観光地に赴いて散策しながら放送を行うということが考えられる．この場合，放送者は全方位カメラを身体に固定したり，スタビライザーなどを用いて手に持ったりすることが考えられる．

放送者が観光地などで移動しながら放送を行った場合，視聴者が興味を持ったものの，放送者が話題にすることなく通り過ぎてしまうというケースが考えられる．視聴者がコメントを用いて指摘したとしても，放送者の視野角にない物体に対する指摘やコメントは放送者に理解されづらい．そこで，放送者の視野角にないものに対する視聴者の興味や関心が向けられた方向を事前に分析する必要がある．視聴者が興味や関心を向けている方向を事前に分析することで，放送者は自身が気づかなかったものについて気づくことができるようになり，視聴者の興味や関心を満たすことができるようになる．本研究では，可搬式全方位カメラを用いた 360 度インターネット生放送において，放送者の視野角にないものに対して向けられた視聴者 POV を検出するアルゴリズムについて検討する．

2. 関連研究

360 度動画や 360 度インターネット生放送における，視聴者の視聴方向分析に関する研究や事例は複数存在する．ここでは，研究の関連する研究や事例を挙げ，本研究の必要性と新規性について述べる．

VRCHEL[4]は，株式会社ジョリーグッドが開発している 360° VR のコンテンツ解析とユーザ行動のパターン解析を人工知能によって行うという取り組みである．視聴中の 360 度動画のリアルタイム解析を動画にオーバーレイ表示することが出来，人物や物体，話されている言語，雰囲気，シーン構成，ストーリーなどの様々な動画の構成要素を認識することが出来る．当該システムはビジネスに活用される VR コンテンツについて，視聴者のニーズ分析やより質の高い 360 度動画の作成を支援することが目的とされている．そのため，当該システムでは，一般ユーザが利用する全方位カメラを用いたインターネット生放送への利用は想

1 岩手県立大学大学院 ソフトウェア情報学研究科
Iwate Prefectural University,
Graduate School of Software and Information Science

ⓒ 2019 Information Processing Society of Japan

定されていない．

Yen-Chen Lin らが行った 360 度動画における視聴方向の補正に関する研究[5]では，視聴者の視聴方向を 360 度動画の主となるストーリーの方向に補正する手法について検討している．当該研究では視聴者の視聴方向を補正するために自動補正機能とアノテーションを用いたガイドの 2 パターンを実装，評価している．その中で，視聴者の視聴行動には複数の目的およびパターンが存在することを明らかにしており，より質の高い視聴体験を提供するためには，視聴者の視聴方向を分析する必要があることを強調している．

360 度動画および 360 度インターネット生放送サービスを提供する YouTube が公開した 360 度動画の分析結果[6]では，視聴者の POV が動画再生中にどこに向けられるのかを分析している．2017 年 6 月に YouTube では提供している 360 度動画サービスにヒートマップを用いたオンデマンド動画解析ツールが実装された．解析ツールを用いた初期調査を実施した結果，動画の正面方向を中心とした 90 度の範囲に視聴者の POV が動画再生時間のうち 75%もの間向けられていることが明らかにされた．また，人気のある 360 度動画でも 360 度全範囲を視聴したユーザは全体の 20%程度であるということも述べられている．

これまでに挙げた研究や事例は 360 度動画において，より質の高い視聴体験を実現するために視聴方向の分析が必要であるということを示している．一方で，オンデマンドの 360 度動画を対象にしており，蓄積されたデータを基に視聴者の視聴対象や視聴方向の変更予測などを実現している．360 度インターネット生放送における利用は想定されておらず，リアルタイムな分析手法を検討することは新規性があると言える．

3. インターネット生放送システム

本研究で用いたインターネット生放送システムは WebRTC および Three.js を用いて実装したものである．これは，2016 年に Google が提供する Chrome ブラウザでの Flashplayer のサポート廃止[7]を受けて，インターネット生放送サービスで HTML5 に対応した新方式が提供されるようになったためである[8][9]．HTML5 に対応した WebRTC では，映像や音声の配信やファイル共有など，リアルタイムコミュニケーションに必要な機能が簡単に利用できる．映像の配信は WebRTC に対応した KurentoMediaServer[10]を利用した．また，360 度動画の再生に関しては，Theta_GL[11]というライブラリを使用した．THETA_GL は RICOH 社が販売する全方位カメラ THETA S で撮影したエクイレクタングラー形式の映像を，360 度動画として Web ブラウザ上で再生することができる．再生には Three.js による WebGL の描画を用いている．再生する映像のソースを WebRTC サーバから取得することでインターネット生放送にも対応できる．図 1 にシステムの構成を，図 2 にシステムのインタフェースを示す．

本システムでは，WebGL の 3D 空間上に球体オブジェクトとカメラを設定している．球体の内側表面には 360 度動画の映像がマッピングされており，球体の中心にカメラ位置が設定されている．視聴者の POV は球表面の極座標系 (ϕ, θ, r)によって管理されている．システムは Web ブラウザを通して利用することができ，視聴者が Web ページにアクセスすると ID が割り振られ，放送の視聴準備が行われる．視聴者クライアントでは，視聴を行っている間，100ms ごとに視聴者の POV を取得し，視聴者の ID と時刻データを付加した上で POV サーバへと送信する．送信された POV データはリスト化され，POV サーバ内に保存される．

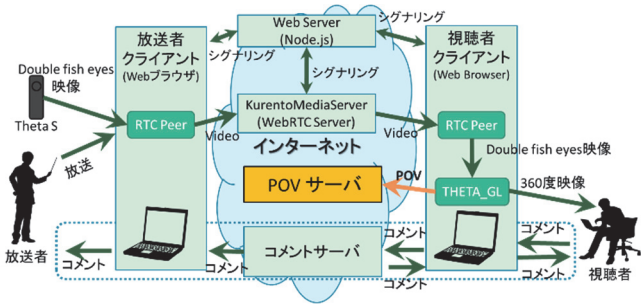

図 1：システムの構成

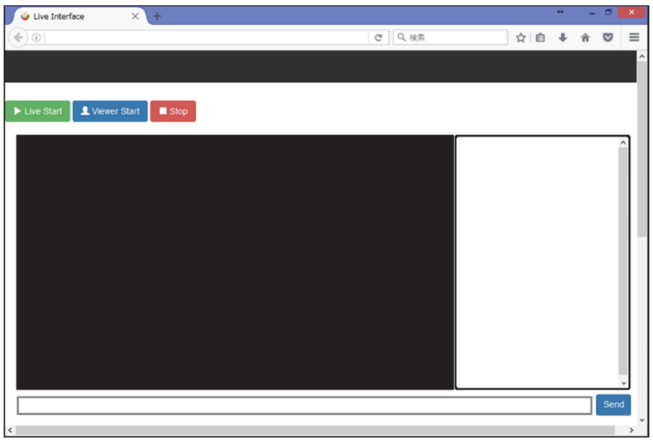

図 2：システムのインタフェース

4. 仮説の設定

我々は関連研究で明らかにされた結果を基に 360 度インターネット生放送における視聴者 POV の動作に関する仮説を立てた．1 つ目の仮説は「視聴者の POV は放送者の進行方向に向けられる」である．オンデマンド動画の場合，編集が加えられており，基準となる動画の正面方向が明示的である．それに対し，360 度インターネット生放送では，リアルタイムな映像であるため編集が不可能で，明示的な動画の正面方向は存在しない．そのため，放送者の進行方向が動画の正面方向として理解されるのではないかと考えた．2 つ目の仮説は「視聴者が興味を持ったものがある場合，その対象に対して視聴者の POV が追従する」である．360 度インターネット生放送において，その特徴である POV の変更は，視聴者が興味関心をもとに行われている．しかし，放送者が移動しながら放送を行った場合，1 度の POV 変更では，興味関心の対象が視聴範囲から外れてしまうことが考えられる．そのため，対象を追従するように複数回の POV 変更が行われると考えた．3 つ目の仮説は「興味の対象が見切れるか，興味を失うと POV は進行方向に向けられる」である．これは 1 つ目の仮説と同様で，動画

視聴の基準となる正面方向が明示的でないため，放送者の進行方向が基準方向として理解されると考えた．

以上のことから 360 度インターネット生放送において，視聴者の POV は基本的に放送者の進行方向に向けられており，興味や関心の対象が見つかると追従するように POV が進行方向から外れる．興味や関心を失うか，対象物が見切れると視聴者の POV は放送者の進行方向に再び向けられると考えられる．

5. 実験

仮説の検証を行うため放送実験を実施した．放送実験の目的は，先に述べた仮説の検証およびアルゴリズム作成に必要となるデータの収集である．今回の放送実験では実験協力者として 6 名が参加し，各回 3 名の協力者が視聴者を担当して 2 回実施した．放送者は 2 回ともに高田が担当した．放送場所は白鳥の飛来地としても知られる岩手県盛岡市高松にある高松ノ池周辺である．池の周囲を右回り左回りそれぞれ 1 回ずつ，30 分の放送時間で行った．放送者の取ったルートについて，1 回目の放送実験を図 3 に，2 回目の放送実験を図 4 に示す．今回，全方位カメラは放送者が背負うリュックに対して一脚を固定して付けた．そのため，映像は放送者の後方上部を中心としたものになっている．視聴者には 360 度インターネット生放送のテスト放送を実施すると伝え，放送を視聴してもらった．実験中は 0.1 秒ごとに視聴者の POV データを取得し，サーバ内に保存している．

放送実験終了後に視聴者の POV データを分析したところ，1 回目の放送では放送者の進行方向を中心とした 90 度の範囲に視聴者の POV が 75.89%の時間向けられており，2 回目の放送でも，同様の範囲に 80.33%の時間向けられているということが確認できた．YouTube が公開されているデータ分析の結果では基準方向を中心とした 90 度の範囲に視聴者の POV が 75%の時間向けられていた．今回収集したデータではこれに近しい値を求めることができたため，収集したデータを基にデータ分析を進めることにした．また，POV が進行方向を中心とした 90 度の範囲から外れた際に，視聴者の興味や関心を惹く対象が POV の方向に存在しているケースが複数確認できた．POV 進行方向を中心とした 90 度の範囲から外れた場合，平均 11.41 秒後に進行方向の範囲に POV が戻ることも確認できた．このことから，仮説が確かであると考えた．

6. アルゴリズムの検討

本研究の目的は，放送者の視野角から外れた対象に興味や関心を持っている視聴者を POV データから検出し，コミュニケーションに活用することである．そのために，視聴者の POV データをもとに追従状態を検出するアルゴリズムを開発する．アルゴリズムを検討するにあたり，放送者の視野角をシステム上で定義することとした．実験において視聴の基準方向は放送者の進行方向になっていることが確認されたため，放送者の視野角も進行方向を基準とすることにした．また，人間の主要視野角は水平方向に 100 度と言われているため[12]，進行方向±50 度を放送者の視野角と定義する．図 5 にエクイレクタングラー方式における放送者の主要視角および周辺視角の模式図を示す．

図 3：実験 1 回目における放送者の進行ルート

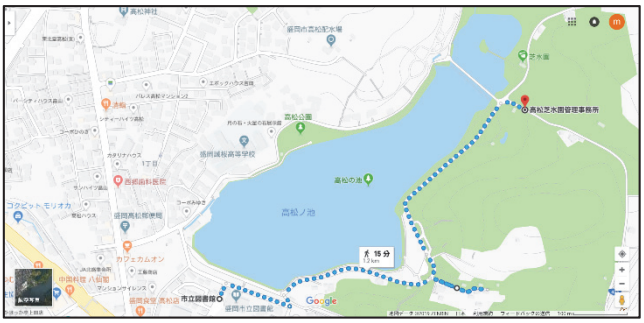

図 4：実験 2 回目における放送者の進行ルート

先の実験において，視聴者の POV を分析した結果，視聴者は自身の興味や関心に合わせて対象物を追従しながら視聴することが確認された．よって，放送者の視野角外で視聴者の POV が追従を行ったことを検知するアルゴリズムを検討する．実験によって得られたデータをもとに，視聴者の状態の分類，および検知に用いるバッファサイズを決定した．視聴者の状態は放送者の進行方向を視聴している「通常状態」，興味関心を惹かれる対象物を発見した「追従開始状態」，興味関心を惹かれる対象物を追従して視聴している「追従状態」，興味関心を失った，または対象物が見切れた「追従終了状態」の 4 つの状態に分類した．通常状態は視聴者の POV が主要視角の範囲内で変化している．追従開始状態は，放送者の主要視角の範囲から周辺視角へと POV が変化する．追従状態は放送者の周辺視角内で POV が変化している．追従終了状態は放送者の周辺視角から主要視角への POV が変化する．ただし，POV データは 0.1 秒毎に取得しているため，10 件分(1 秒分)のデータをまとめてアルゴリズムで使用することとした．

以上のことを踏まえて作成したアルゴリズムのフローチャートを図 6 に示す．フローチャートで使用する変数と条件式について，図 7 に示す．作成したアルゴリズムではまず，進行方向を 0～360 の水平方向の角度座標として取得する．その後，最新の POV データから 10 件分のデータをもとに 4 つの状態のうち，どれに該当するのかを条件式によって判断する．誤検出の割合を考慮し，同じ状態が 10 回，つまり 1 秒間継続した場合，状態を遷移することとした．今回，1 秒間の継続で状態遷移すると設定したのはリアルタイム性の担保と誤検出の防止を検討した結果である．

これにより，主要視角と周辺視角の境界線付近を視聴している際に，周辺視角側に出たとしてもすぐ戻った場合には追従検知から外れることとなり，誤検出を抑止することができる．一方で視聴者が追従を開始してから 1 秒で追従を検知可能になるため，即座に視聴者の追従対象について確認することができる．

また，放送者の進行方向について，全方位カメラの特性上，設置および固定がどのように行われるかが確定的でない．そのため，放送のたびに進行方向がずれてしまうことが予測される．本アルゴリズムを実装する際には，全方位カメラに搭載された加速度センサなどを用いて進行方向を取得，補正する必要が考えられる．

7. アルゴリズムの適用と分析

作成したアルゴリズムを放送実験で収集したデータに対して適用し，放送者の視野角外での追従を正しく検出できているか, 追従の対象が何なのかについて分析を実施した．その結果，1 回目の放送では 73 件，2 回目の放送では 95 件，合計 168 件の追従を検出することができた．

さらに，検出された追従時のデータについて，追従の対象と放送における話題への利用の観点から，3 段階に分類することにした．追従の対象が明らかで，生放送で話題にしやすいものを A，追従の対象が明らかではあるが生放送で話題にするのが困難なものを B，追従の対象が明らかでない，または追従ではないものを C として分類する．追従として検出された 168 件のデータのうち A に分類されたものが 76 件(45.23%)，B に分類されたものが 73 件(43.45%)，C に分類されたものが 19 件(11.30%)となった．各放送における 3 分類の件数とその追従対象を表 1 に示す．

A として分類されたものでは，池のほとりに建てられた小さな社や，飛来した鴨の着水などを追従しているものが見られた．B として分類されたものでは，池の外周に停められた自動車に対して注目したものや，主要視角と周辺視角の境界線付近にいる鳥の群れに注目したものなどが見られた．C と分類されたものとしては，放送を開始した直後や開けた場所に着いた際に周囲を見渡すという場合や, 360 度インターネット生放送に慣れていなかったため動作確認を行った場合などが見られた．

追従が検出されたデータについて，追従が継続された時間の長さについて確認したところ，A は 9.76 秒，B は 12.09 秒，C は 15.36 秒となった．標準偏差については，A が 7.89，B が 23.37，C が 23.27 となった．ただし，B と C については追従の継続時間が 100 秒を超えるデータが存在しており，極端な外れ値が含まれていると考えられる．これらの外れ値を取り除いた場合の継続時間は A が 9.76 秒，B が 9.73 秒，C が 10.44 秒となり，それぞれの追従継続時間について大きな違いは見られなかった．標準偏差は A が 7.89，B が 12.13，C が 10.54 となった．外れ値として処理したデータは，周辺視野角の範囲で 100 秒以上移動を停止していたことが分かった．これは視聴者が POV の移動中にコメント入力操作を行ったために発生したことが分かった．この結果から，視聴者の追従動作は 10 秒前後で終了することが分かった．

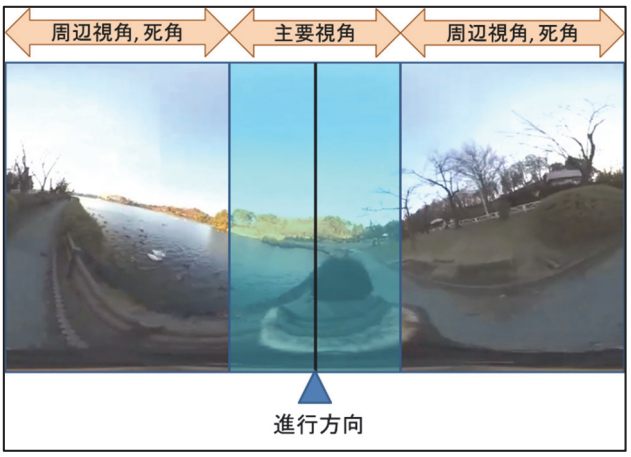

図 5：放送者の視角の模式図

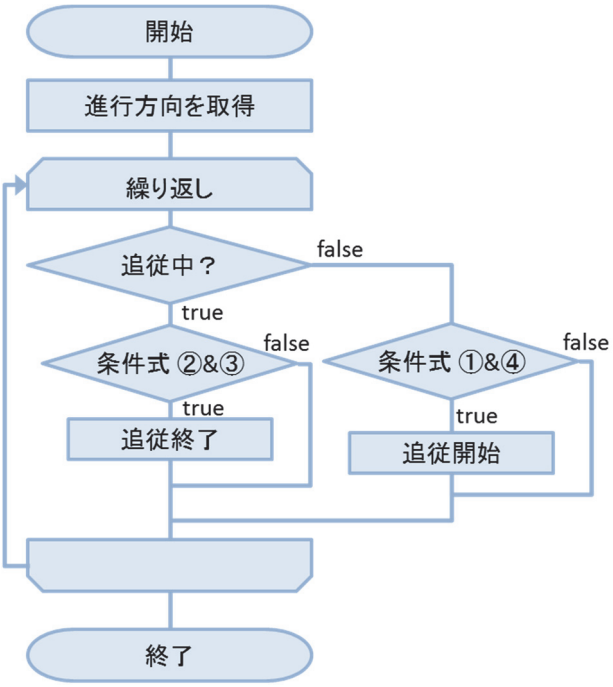

図 6：作成したフローチャート

変数

$\emptyset_i = \{\emptyset_{i1}, \emptyset_{i2} \ \dots \ \emptyset_{in}\}$ 視聴者 POV の水平方向極座標

$\theta_i = \{\theta_{i1}, \theta_{i2} \ \dots \ \theta_{in}\}$ 視聴者 POV の垂直方向極座標

forward = [進行方向の∅座標]

条件式①

forward - 50 < ∅_(i n−9) < forward + 50

条件式②

∅_(i n−9) < forward - 50 || forward + 50 < ∅_(i n−9)

条件式③

forward - 50 < ∅_(i n) < forward + 50

条件式④

∅_(i n) < forward - 50 || forward + 50 < ∅_(i n)

図 7：各変数と条件式

8. 考察

今回作成したアルゴリズムでは，放送者の主要視野角外の対象に向けられた視聴者の POV 追従を検出した．アルゴリズムによって検出されたデータのうち，追従対象が推測可能なものは 168 件中 149 件で, 88.69%のデータである．ただし，コミュニケーションに活用できるかを考慮した場合，有用なのは 168 件中 76 件の 45.23%に留まった．

追従対象の推測ができなかった 19 件のデータに関しては，周囲の状況を確認するための見渡し行為や 360 度インターネット生放送の動作確認など，追従動作に近い動きをしているものを誤検出した場合があった．また，追従対象は推測できるものの，コミュニケーションに生かしづらいと判断されたものでは，主要視角と周辺視角の境界線付近を視聴していて検出されたものが多かった．今回設定した進行方向±50 度という主要視角の範囲に関して再検討する必要がある．

今回検出された追従について，継続時間が約 10 秒であることが分かった．このことから，追従の検出に用いることができるデータ量は 100 件程度ということが分かった．今後は検出率および誤検出の割合を基に適切なアルゴリズムで用いるデータ長を決定する．

9. 今後の展望

今回アルゴリズムの作成を行った結果，追従対象が推測可能だが話題性に乏しいものが全体の 43.45%存在した．今回の分析において，放送者が話題性についての判断を行ったため，追従対象を生放送での話題に利用できるか，という点に関する明確な基準を設定することができなかった．話題性の有無に関する基準の設定は今後の課題として検討する必要がある．その上で，POV から話題性の有無についての判断が可能かどうかについて検討を進める．

今回データを分析する中で誤検出されたものは，①主要視角と周辺視角の境界線付近を視聴しており，ごく短い時間周辺視覚側へ POV が移動したもの, ②追従の途中で POV の動作を停止させたもの，③見回し動作などの想定外の動作を行ったものの 3 パターンが存在した．それぞれのパターンへの対策を検討する必要がある．①への対策は主要視角と周辺視角の境界線設定を変更することである．今回設定した視線方向±50 度の主要視角範囲は，顔や眼球を動かさない場合を想定したものであり，顔の向きが固定されていない放送者を対象とした場合，主要視角の範囲を広く取る必要があると考えられる．どの程度の範囲を主要視角とするべきかについては今後調査をする必要がある．また，主要視角と周辺視角の境界線からの距離に基づいてデータの重み付けを行うという方法も考えられる．②への対策としては，視聴者の POV 移動量を求めて，フィルターをかける方法が考えられる．③への対策としては，検出に用いる POV データをベクトルデータに変換し，内積を求めることで蛇行の程度を求めることで追従対象が存在するか否かを分析することができるのではないかと考えられる．今後はアルゴリズムの修正を行った上で，検出された追従の方向を放送者に提示する方法について検討を行う．

表 1：各放送における追従の分類と追従対象

放送	A 追従対象○ 話題性○	B 追従対象○ 話題性×	C 追従対象×
放送 1回目	35件	29件	9件
	手漕ぎボート 池のほとりの社 特徴的な建物	すれ違った通行人 遠ざかる鴨の群れ 対岸にいる白鳥	見回し 動作確認
放送 2回目	41件	44件	10件
	鴨の着水 ビニールハウス 池を撮影する人	歩道脇の樹木 駐車場の自動車 夕日	見回し 無意味な動作

10. まとめ

本研究では，360 度インターネット生放送において視聴者が放送者の視角外の物体に対して追従して視聴していることを検出し，コミュニケーション支援に活用する方法を検討した．その結果，視聴者の POV を用いた追従検出アルゴリズムを作成し提案した．アルゴリズムを用いてデータ分析を行ったところ，追従対象の推定が可能な検出データが全体の 88.69%になった．ただし，コミュニケーションへの活用が可能なデータは全体の 45.23%になり，追従検出のアルゴリズムに改良を加える必要がある．

参考文献

[1] Roel Vertegaal, “The GAZE groupware system: mediating joint attention in multiparty communication and collaboration”, CHI '99 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp.294-301(1999).

[2] David M. Grayson, Andrew F. Monk, “Are you looking at me? Eye contact and desktop video conferencing”, ACM Transactions on Computer-Human Interaction (TOCHI) Volume 10 Issue 3, September 2003, pp.221-243(2003).

[3] 高田真也，西岡大，齊藤義仰：全方位カメラを用いたインターネット生放送における視線ヒートマップの開発，分散協調とモバイルシンポジウム 2017 論文集，pp.1081-1088(2017).

[4] VRCHEL, https://vrchel.com/, (2017 年 12 月参照).

[5] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, Min Sun, “Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video”, CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 2535-2545(2017).

[6] 360° 動画のヒートマップレポート, https://support.google.com/youtube/answer/7407544/, (2017 年 12 月参照).

[7] Intent to implement: HTML5 by Default, https://groups.google.com/a/chromium.org/forum/#!searchin/chromium-dev/HTML5$20by$20default/chromium-dev/0wWoRRhTA_E/__E3jf40OAAJ, (2017 年 12 月

参照).
[8] YouTube Engineering and Developers Blog: YouTube now defaults to HTML5 <video>, https://youtube-eng.googleblog.com/2015/01/youtube-now-defaults-to-html5_27.html, (2017 年 12 月参照).
[9] 「動画視聴ページ HTML5 版（β）」提供開始のお知らせ - ニコニコインフォ, http://blog.nicovideo.jp/niconews/ni064261.html, (2017 年 12 月参照).
[10] Kurento, http://www.kurento.org/, (2017 年 12 月参照).
[11] Theta_GL, 2016, https://github.com/mganeko/THETA_GL, (2017 年 12 月参照).
[12] 早田赳, 岩切宗利, “周辺視野への視角外空間の表示による視野拡張に関する一検討”, 研究報告グラフィクスと CAD(CG) 2015-CG-158(10), pp.1-5(2015).

ⓒ 2019 Information Processing Society of Japan