Vol.2018-AVM-103 No.19
2018/11/30

360 度映像視聴時の視野移動履歴と映像の顕著性に関する分析評価

篠原裕矢　白崎智美　呉益妍　金井謙治　甲藤二郎
早稲田大学基幹理工学部 〒169-8555 東京都新宿区大久保 3-4-1 55N-6F-09
E-mail: {y_shinohara, kanai, katto}@katto.comm.waseda.ac.jp

概要：近年，VR/AR/MR といった XR コンテンツの普及に伴い，高品質な 360 度映像配信の需要も高まっている．360 度映像は，コンテンツが高解像度なだけでなく，視野角も広くなるため，より大容量になりやすい．そのため，低通信量で品質を保つような効率的な配信手法が求められている． ユーザーの視野移動を予測し，視野領域を利用して配信を制御することができれば，効率的な配信手法につながることが予想される．しかし，映像コンテンツごとにユーザーの視野移動の特徴は変わってくる．そこで，本稿では，360 度映像視聴時の視野移動の履歴および映像の顕著性について，コンテンツドリブンに分析し，それぞれの映像コンテンツにおける適切な予測方法について検討する．

キーワード：360 度映像配信，顕著性マップ，適応レート制御，ROI 符号化

Performance analysis of viewport movements and saliency maps in 360-degree video experiences

YUYA SHINOHARA　SATOMI SHIRASAKI　YIYAN WU　KENJI KANAI
JIRO KATTO
Fundamental Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjyuku-ku, Tokyo, 169-8555 Japan
E-mail: {y_shinohara, kanai, katto}@katto.comm.waseda.ac.jp

Abstract: Recently, XR contents such as VR / AR / MR are widely spreading, and the demand for high-quality 360-degree video delivery is also increasing. The 360-degree video requires a larger capacity because the content needs high resolution and its viewing angle is wider. Therefore, an efficient delivery method, such as reducing network traffic while keeping high visual quality, is required. Prediction of the movement of the user's field of view (FoV) will provide an efficient rate (and quality) control method. However, characteristics of the FoV movement may change according to video contents and users. Therefore, in this paper, we analyze historical FoV movements and the saliency of 360-degree videos and provide an appropriate prediction method for each video content.

Keywords: 360-degree video delivery, Saliency Map, Adaptive rate control, Region of Interest coding

1. はじめに

近年，Virtual Reality (VR) / Augmented Reality (AR) /Mixed Reality (MR)などの XR コンテンツの普及が急速に進み，それに伴って高品質な 360 度映像配信の需要も高まっている．2020 年までにそれらの市場規模は 2017 年のおよそ 6 倍に達すると予想されている[1]．また，2016 年は VR 元年と呼ばれ，HTC Vive[2]や Oculus Rift[3]，PlayStation VR[4]など様々なヘッドマウントディスプレイ(HMD)が比較的安価で提供されているため，研究用途のみならずコンシューマ用途にも普及し始めている．さらに 2020 年の東京オリンピックに向けて，360 度映像によるリアルタイム配信も検討されており，今後は 360 度映像の放送も普及していくことが予想される．

一方で 360 度コンテンツは，4K/8K といった映像の高解像度化の流れに加え，広い視野角に対応することも求められているため，コンテンツの大容量化が懸念されている．そのため，360 度コンテンツの配信では，高品質かつ低通信量といった，より効率的な適応レート制御手法が求められている．

筆者らはこれまで，[5]を参考にして，360 度映像配信時のユーザーの視野領域を考慮した適応レート制御を行うことで，通信量を抑えた高品質な 360 映像配信の検討を行ってきた[6]．本稿では，課題の一つとして残っていたユーザーの視聴時の視野予測について検討する．具体的には，機械学習手法の一つであるサポートベクターマシン（SVM）を用いて，360 度映像視聴時に，ユーザーが映像のどの領域を見るのかを予測し，その精度を評価する．

360 度映像視聴時にユーザーが注目しやすい領域の要因として，過去の視野移動の履歴だけではなく，コンテンツごとの特性も関係していると思われる．そこで本稿では，特徴量として映像の顕著性に注目する．顕著性とは，人の

ⓒ2018 Information Processing Society of Japan

注視の引きつけやすさを示す指標のことである．フレームの各ピクセルに対して注視の度合いを推定したものを顕著性マップ[16]と呼び，今回はこれを利用する．映像の顕著性を用いて，複数の種類の 360 度コンテンツをクラスタリング手法によって分析し，グループごとに適切な予測モデルを切り替えることで予測精度を上げることも検討する．

本実験では，実際に HMD を用いてユーザーに 360 度映像を視聴してもらい，ユーザーの視野領域のトラッキングデータを収集し予測する．精度評価では，映像をいくつかのタイルに分割し，視野領域が予測したタイルに入っているかどうかの精度（Precision）を評価する．

2. 関連技術および関連研究

2.1 視野領域を考慮した 360 度映像の MPEG-DASH 配信

MPEG-DASH は，近年の映像配信技術において主流となっており，ネットワークの通信状況に応じて映像コンテンツの配信ビットレートを動的に変更する「適応レート制御」採用している．MPEG-DASH では，1 つのコンテンツを数秒のセグメントごとに分割し，各セグメントに対して，解像度，ビットレートを複数用意することで，コンテンツの階層化を実現している．各階層はレプリゼンテーションと定義している．これらは，Media Presentation Description (MPD)という XML 形式のファイルで管理しており，このファイルをセッション開始時に取得，参照することで，通信帯域に応じて最適なレプリゼンテーションを選択することを可能としている．

360 度映像の DASH 配信拡張として，Region of Interest (ROI)符号化を適用している手法が提案されている[5, 14]．[5]では，ユーザーの視野領域も考慮して，DASH コンテンツの階層化を行っている．そのため，ネットワーク品質だけでなく，ユーザーの視野領域にも応じて，レプリゼンテーションを適応制御することで，配信品質の向上を実現している．[14]では，この視野領域の定義として，画面全体を細かなタイルに分割し，視野領域に応じて，各タイルの符号化率を適応変化させ，より柔軟な品質制御を実現している．

2.2 サポートベクターマシン(SVM)

SVM は，パターン識別用の教師あり機械学習方法の一つである．「マージン最大化」というアルゴリズム等で汎用性も高く，現在知られている方法としては，最も優秀な学習器の一つとして知られている．図 1 にそのイメージを示す．点の集合に対して境界となるような線を引くことで識別するアルゴリズムである．正しい分類基準を見つけるために，「マージン最大化」を用いており，これにより少しだけデータが変わった場合の誤判定などを防ぐことができる．

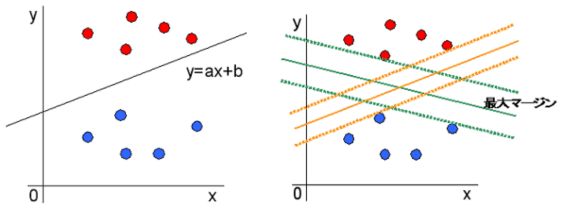

図 1. SVM のイメージ[15].

2.3 360 度コンテンツ投影手法

360 度映像は，一意のポイントからあらゆる方向にキャプチャされたものであるため，本質的には球形の映像となっている．しかし実際にサーバーで保管する際は，通常の動画と同じように矩形の映像として保存されている．その映像を HMD といったクライアントで表示する際，幾何学変換を行っている．このような，球面から平面へ（あるいは逆）の幾何学変換の一つとして，図 2 のような正矩円筒図法（Equirectangular）が知られている．

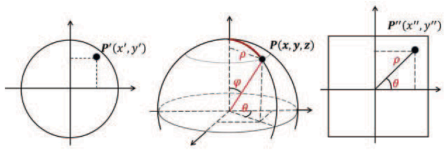

図 2. 球面から平面への変換の流れ[9].

図 2 に示されている球面上の黒点 $P(x, y, z)$は，右の図の平面上の黒点 $P''(x'', y'')$にマッピングされる．その際の変換式は(1)式，(2)式のように計算される．

$$x'' = 0.5W + \frac{H}{f}\tan^{-1}\frac{\sqrt{(\sin\varphi\sin\theta)^2+\cos\varphi^2}}{\sin\varphi\cos\theta}\cos\theta \quad (1)$$

$$y'' = 0.5H + \frac{H}{f}\tan^{-1}\frac{\sqrt{(\sin\varphi\sin\theta)^2+\cos\varphi^2}}{\sin\varphi\cos\theta}\sin\theta \quad (2)$$

ここで，W は平面座標にしたときの幅，H は高さである．また，f は焦点距離，φとθは図 2 の球面で示されている角度である．この例では，正距円筒図を紹介したが，他にもピラミッド型やキューブマップ型といった様々な方式の幾何学投影の方法が知られている．

3. 360 度コンテンツの分析と視野予測

本実験では，360 度映像を視聴するため，HMD の一つである HTC Vive Pro を用いる．また HMD のセンサ情報を用いて，視野移動の履歴を角度で記録する．その角度から 2.3 節の投影手法を用いてタイル位置に変換することで視野移動予測を行う．予測手法には scikit learn[18]で実装されている SVM の一つであるサポートベクター分類（SVC）を用いる．

まず，本稿で用いるデータセットについて説明する．360 度映像配信実験では，合計 138 名の被験者に 1~3 種類ずつ

の 360 度映像を視聴してもらった．その際，被験者には HMD の視野移動に加え，脳波を測定できる MindWave [10]，脈拍を測定できる Fitbit [11]も同時に装着し，映像視聴中の各情報を収集する．用意した 360 度映像は 8 種類でその特徴は表 1 に示す通りである．実験を通して，のべ 275 回分の視聴データを収集した．なお，各映像の視聴時間は 60 秒とし，解像度は 4K である．

表 1. 360 度映像コンテンツの特徴.

動画名	コンテンツ	視聴数
basket	バスケットボールの試合映像	15 回
horror	殺人鬼の出てくるホラー映像	64 回
mountain	山を上空から撮影した風景映像	17 回
paris	パリの街の風景映像	9 回
roller	ジェットコースターからの映像	79 回
shark	海の中でサメが泳いでいる映像	44 回
slider	スライダーから滑り降りる映像	39 回
soccer	サッカーの試合観戦映像	8 回

実験中の視野領域の収集には opentrack [17] を用いており，ユーザーの頭部の視聴方向の角度の収集を行った．これらを 2.3 節で紹介した幾何学変換で正距円筒図法形式に変換する(図 3)．opentrack では Yaw 角，Pitch 角，Roll 角，x 座標，y 座標，z 座標の情報を取得できる．360 度映像を視聴している際の頭部の横振りは Yaw 角（±180 度），縦振りは Pitch 角（±90 度）で表される．今回は単純化のために，Yaw 角と Pitch 角の 2 つを用いて視野領域（タイル）を算出している．

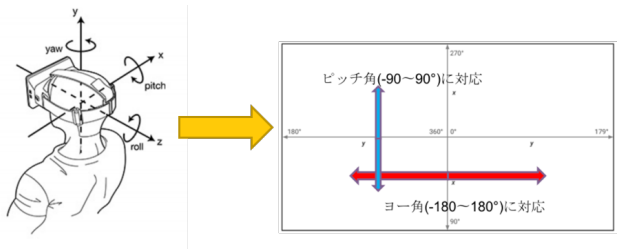

図 3. トラッキング結果の幾何学変換.

3.1 SVC 予測モデルの概要

筆者らは[12]にて，360 度映像視聴時の視野移動の角度履歴を用いたサポートベクター回帰（SVR）による視野移動の予測を行ってきた．本稿では，視野移動の角度履歴だけではなく，コンテンツの顕著性[16]も組み込んだ予測モデルを提案する．具体的には，正距円筒図の形式になっている映像を 64 個（縦 8 個×横 8 個）のタイルに分割し，HTC Vive Pro よって取得される角度から視野がどのタイルに該当するのかを算出する．

視野タイル位置の移動軌跡の履歴とそれらのタイルの顕著性を入力とした SVC による視野移動の予測を行う．モデルの概要を図 4 に示す．

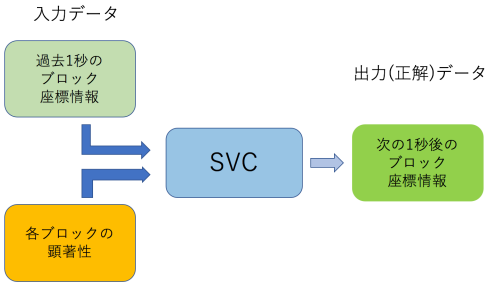

図 4. SVC による予測モデルの概要.

データの入力としては，1 秒前の視野移動履歴を入れて学習させ，1 秒先の視野を予測することとする．動画ごとにタイル位置が一致すれば正解とみなして Precision（精度）を評価する．データの 75%を訓練データに，残りの 25%をテストデータに用い，学習は表 1 に示されている異なる種類の映像に対して，独立して行うこととする．評価指標には Precision（精度）を用いる．

また，顕著性を考慮した場合と考慮しなかった場合の精度比較も行い，また各特徴量の予測結果に対する重要度分析も行う．

3.2 映像コンテンツのクラスタリング分析

予測を行う際，各映像コンテンツを独立して学習する場合，似た特徴を持つ映像コンテンツのみを抽出して学習する場合，すべての映像コンテンツを利用して学習する場合の 3 種類の学習法を検討する．似た特徴を持つ映像コンテンツを抽出する際，事前にクラスタリング分析を適用することとする．クラスタリングとは，分類対象の集合に対して，特徴量分析をすることで，部分集合に分割する手法である．今回の実験に用いた映像は 8 種類あるが，被験者が「よく視野を動かす映像」,「あまり視野を動かさない映像」などといった，何かしらクラスタリングできる特徴があるのではないかと推測できる．そこで，視野移動の履歴をもとに教師なしのクラスタリング分析を行う．クラスタリングには階層的手法と非階層的手法があるが，今回は各映像のサンプル数に差があるため，階層的手法の一つであるウォード法[13] を用いて分類を行うこととする．図 5 に今回用いた階層的クラスタリング（ウォード法）の概要図を示す．このようにデータセットを効率的に学習することで，学習データ量の増加やノイズ除去など，予測精度が向上することが期待される．

ⓒ2018 Information Processing Society of Japan

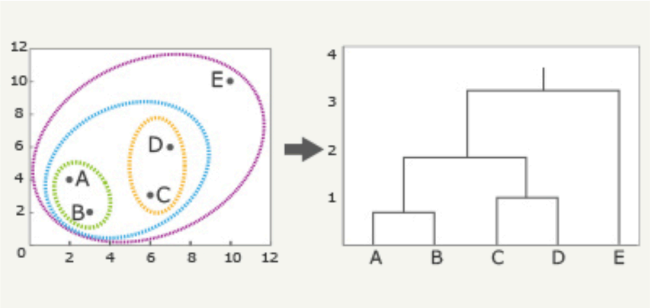

図 5. 階層的クラスタリングのイメージ[13].

4. 評価実験

4.1 SVC を用いた映像コンテンツごとの 360 度映像視聴時の視野予測

まず本実験では，3.1 節で述べた SVC 予測モデルを用いて，精度評価を行った．その際，8 種類の映像コンテンツをすべて独立して学習させて精度を求めている．

実験結果を図 6 に示す．横軸はコンテンツごとに並んでおり，緑色の棒グラフが顕著性なしの場合，青色の棒グラフが顕著性ありの場合である．縦軸が Precision である．図 6 を見ると，全体としておおよそ 75%程度以上の精度を達成している．

また，コンテンツ別に顕著性の影響を見ていくと，basket, roller，slider の 3 つのコンテンツで顕著性ありの予測のほうが高い精度になり，残りの mountain，paris，shark，soccer の 4 つのコンテンツで顕著性なしの予測のほうが高い精度になった．horror のコンテンツは顕著性の有無であまり変わらないという結果となった．この原因は映像のコンテンツの特徴にあると考えられる. 例えば paris のような風景映像は，視聴者によって様々な場所を見渡しやすいため，顕著性がノイズとなっている可能性があると推測される．それとは逆に，roller のような乗り物に乗っている映像は，比較的正面にある顕著性の高いものが注視されやすく，顕著性が説明変数としてうまく機能していると推測される．そのためこの結果から，顕著性を考慮するかどうかはコンテンツによって変わってくるということが言える．

4.2 特徴量の重要度分析

4.1 節の結果を踏まえ，特徴量の重要度分析を行った．重要度とは，予測結果に対して各特徴量がどれだけ寄与したかを表す相対的な値である．本実験における特徴量とは，視野移動の履歴と顕著性である．特徴量の重要度分析の結果を表 2 に示す．比率は視野移動が顕著性の何倍の影響度であるかを表している．

表 2. 特徴量の重要度分析結果.

コンテンツ	視野移動(A)	顕著性(B)	比率(A÷B)
basket	0.2518	0.0181	13.91
horror	0.2957	0.0094	31.46
mountain	0.2618	0.0091	28.77
paris	0.2422	0.0114	21.25
roller	0.3019	0.0139	21.72
shark	0.2350	0.0220	10.68
slider	0.2523	0.0211	11.96
soccer	0.2138	0.0130	16.45

表 2 の結果から，すべての 360 度コンテンツにおいて，顕著性よりも視野移動履歴の方が，10 倍以上予測結果に影響を与えていることがわかる．そのため，顕著性はわずかに予測精度を上げてくれる可能性はあるが，360 度映像の場合は直前にどこを見ていたのかが最も重要なパラメーターであると言える．

ⓒ2018 Information Processing Society of Japan

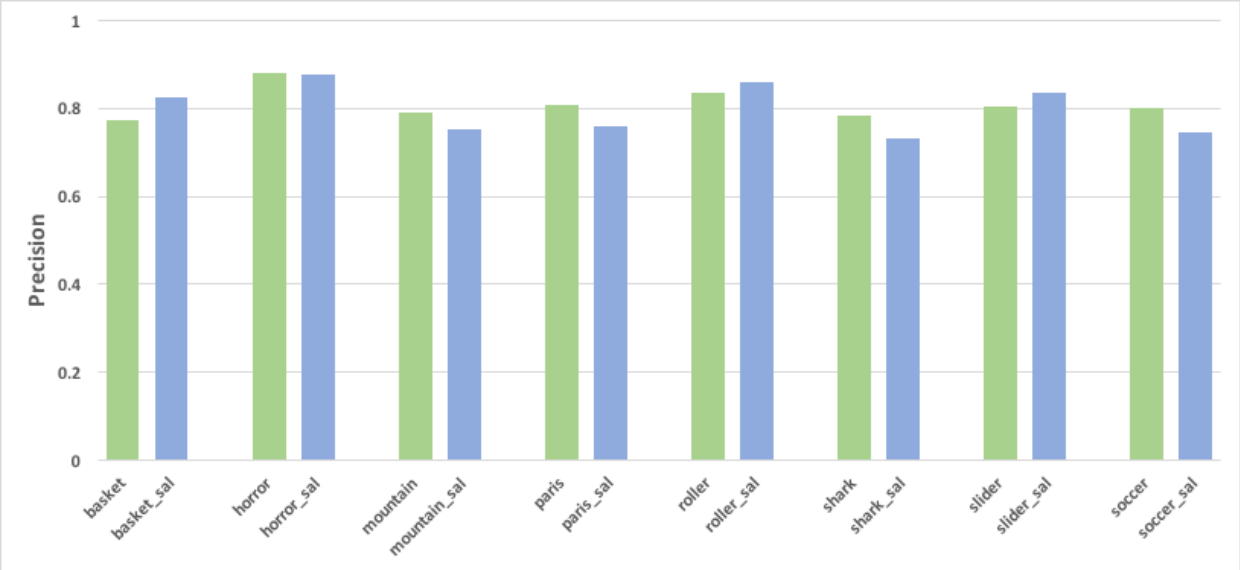

図 6. コンテンツごとの SVC による視野予測の精度結果.

4.3 階層的クラスタリング分析

次に，本実験では，全視聴データ（合計 275 個）の視野移動履歴に対して，クラスタ間の最大距離の 70%を閾値としたウォード法による階層的クラスタリングを行った．分類の結果，class1 と class2 の 2 つのクラスに分かれた．各コンテンツがどのくらいの割合で分類されたかの結果を表 3 および図 7 に示す．

表 3. 各コンテンツの分類割合.

コンテンツ	class1	class2
basket	100.0%	0.0%
horror	98.4%	1.6%
mountain	82.4%	17.6%
paris	77.8%	22.2%
roller	89.2%	10.8%
shark	54.5%	45.5%
slider	94.9%	5.1%
soccer	25.0%	75.0%

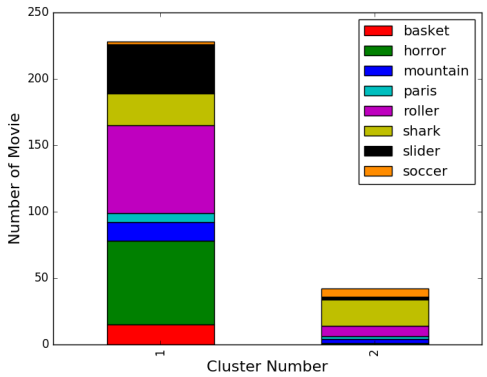

図 7. 階層的クラスタリングの分類結果.

図 7 より，class1 のほうが class2 よりもデータ数がかなり多くなったことがわかる．また，表 3 からコンテンツごとの分類割合を見ると，paris，shark，soccer の 3 つのコンテンツが，他に比べても class2 に属するデータの割合が多かった．これらのコンテンツは 4.1 節の結果において，顕著性を入れたことにより予測精度が下がったコンテンツ群であったため，何らかの関係があると推測される．

続いて，これらの各クラスに対して，4.1 節と同じ SVC を用いた視野予測の精度評価を行った．単純に動画の種類ではなく，階層的クラスタリングの結果で分かれた 2 種類の動画群（class1 と class2）に対して，SVC による視野予測を行いことで，より適切に予測モデルを切り替えられると期待される．

視野予測の精度結果を表 4 および図 8 に示す．横軸はクラスごとに並んでおり，class1+2 は今回の全データをまとめて SVC にかけていることを表している．オレンジ色の棒グラフが顕著性なしの場合，水色の棒グラフが顕著性ありの場合である．縦軸が Precision である．

表 4. クラスごとのデータ数と SVC による視野予測の精度結果.

	データ数	顕著性なし	顕著性あり
class1+2	275 個	0.8419	0.8431
class1	228 個	0.8566	0.8736
class2	47 個	0.6938	0.6509

ⓒ2018 Information Processing Society of Japan

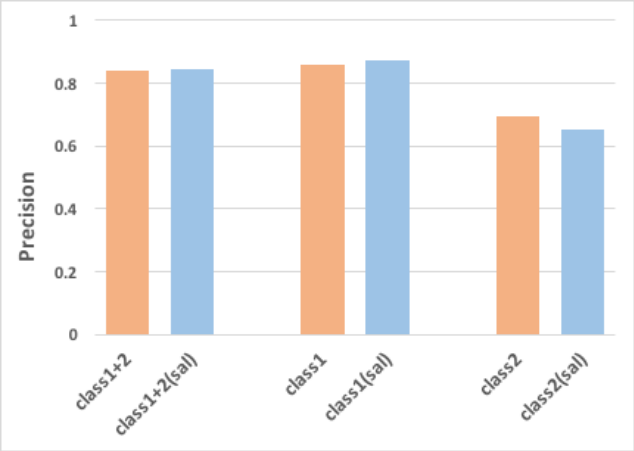

図 8. クラスごとの SVC による視野予測の精度結果.

結果を見ると，class1+2（全データ）と class1 での視野予測では精度が 85%近くあるのに対し，class2 は 70%に届かなかった．class2 のみで視野移動予測を行ったときの予測精度が低くなった理由として，単純にデータ数が少なかったことも挙げられるが, 4.1 節の実験でも映像によってはデータ数が 50 個以下のものもあった. それに比べてもさらに精度が低くなった理由は, class2 に属するデータ群は「視野がよく動くもの」が多かったからだと推測する. class2 に属するコンテンツのうち割合の高かった映像は, 4.1 節の結果で顕著性を入れたことで予測精度が下がったコンテンツと同じ映像が多かったためである.

また，class1+2（全データ）で予測するよりも class1 のデータで予測したほうが，わずかに精度が良くなっている．このことから，データ数が多いことはもちろん大切だが，class2 のような「視野がよく動くもの」をデータに入れても，外れ値やノイズのように働いてしまい，逆に精度が下がってしまうことが分かった．

最後に顕著性の効果に関して考察すると, class1 と class2 ではっきりと分かれた．class1 では 2%ほど予測精度が向上したのに対し，class2 では 4%以上も予測精度が下がった．前の実験と同じく，顕著性がノイズとなっていることが考えられる．

そのため, 本実験条件下に関して言えば, class1 に属するようなコンテンツの予測モデルには「視野移動の履歴+顕著性」を学習させるのが適していて, class2 に属するようなコンテンツの予測モデルには「視野移動の履歴のみ」を学習させるのが適していると考察できる．

5. まとめと今後の予定

本稿では，360 度映像の視聴データを用いた視野移動予測の精度評価を行った. 機械学習手法の一つである SVM を用いることで，本実験条件下では，おおよそ 70～80%程度の精度で視野領域のタイルを予測することができた. また，コンテンツによっては，顕著性も被験者の視野移動に影響を与えることがわかった．

今後は，これらの視野移動予測結果を用いて, 360 度映像配信に対して，リアルタイムで動画像符号化処理をすることや, その際の品質評価を行っていくことを目指す. また，引き続き 360 度映像の視聴データのサンプル数を増やすことも課題である．さらに，より多彩な評価指標を用いることで，360 度映像配信の性能を検証していく予定である．

謝辞 本研究成果は JSPS 科研費 15H01684 および 17K12681 の支援を受けている．

参考文献

[1] Augmented/Virtual Reality to hit $150 billion disrupting mobile by 2020 | NEWS >> Digi-Capital [online]: http://www.digi-capital.com/news/2015/04/augmentedvirtual-reality-to-hit-150-billion-disrupting-mobile-by-2020/
[2] VIVE™ 日本 | 想像を超えたバーチャルリアリティの体験 [online]: https://www.vive.com/jp/
[3] Oculus Rift [online]: https://www.oculus.com/rift/
[4] PlayStation VR [online]: http://www.jp.playstation.com/psvr/
[5] X. Corbillon, et al. “Viewport-adaptive navigable 360-degree video delivery,” IEEE ICC, May 2017.
[6] 篠原，金井，甲藤，“視野領域を考慮した 360 度映像配信の適応レート制御手法の性能評価”, AVM 情報処理研究発表会, 2018 年 3 月
[7] FFmpeg [online]: https://ffmpeg.org/ffmpeg.html
[8] Shanghai-based VR Headset “eyeForce” Starts Kickstarter Campaign [online]: https://medium.com/vr-today-magazine/shanghai-based-vr-headset-eyeforce-starts-kickstarter-campaign-2bdabe81aa26
[9] Tuan Ho, Madhukar Budagavi, “DUAL±FISHEYE LENS STITCHING FOR 360±DEGREE IMAGING”, IEEE ICASSP, March 2017.
[10] MindWave [online]: https://store.neurosky.com/pages/mindwave
[11] アクティビティ記録デバイスなどの情報が得られる Fitbit オフィシャルサイト [online]: https://www.fitbit.com/jp/home
[12] 篠原，金井，甲藤，“SVR による 360 度映像視聴時の視野移動予測の精度評価,” 電子情報通信学会ソサイエティ大会, 2018 年 9 月.
[13] クラスタ分析の手法②（階層クラスタ分析），データ分析基礎知識 [online]: https://www.albert2005.co.jp/knowledge/data_mining/cluster/hierarchical_clustering
[14] Duc V. Nguyen, et al. “A New Adaptation Approach for Viewport-adaptive 360-degree Video Streaming” 2017 IEEE ISM, Dec 2017.
[15] サポートベクターマシン（SVM） [online]: http://www.sist.ac.jp/~kanakubo/research/neuro/supportvectormachine.html
[16] “Saliency Map Algorithm: MATLAB Source Code”,[online]: http://www.klab.caltech.edu/~harel/share/gbvs.php.
[17] opentrack/opentrack: Head tracking software for MS Windows, Linux, and Apple OSX [online]: https://github.com/opentrack/opentrack
[18] scikit learn [online]: http://scikit-learn.org/stable/

ⓒ2018 Information Processing Society of Japan